

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
19 December 2002 (19.12.2002)

PCT

(10) International Publication Number
WO 02/101044 A2

(51) International Patent Classification: **C12N 15/12**,
5/10, C07K 14/47, A61K 38/17

(21) International Application Number: PCT/EP02/06316

(22) International Filing Date: 6 June 2002 (06.06.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
01202214.1 11 June 2001 (11.06.2001) EP

(71) Applicant (for all designated States except US):
JANSSEN PHARMACEUTICA N.V. [BE/BE]; Patent
Department, Turnhoutseweg 30, B-2340 Beerse (BE).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **DEL-FAVERO**,
Jurgen, Peter, Lode [BE/BE]; Flanders Interuniversity
Institute for Biotechnology (VIB), Rijvisschestraat 120,
B-9052 Zwijnaarde (BE). **VAN BROECKHOVEN**,
Christine [BE/BE]; Flanders Interuniversity Institute
for Biotechnology (VIB), Rijvisschestraat 120, B-9052
Zwijnaarde (BE).

(74) Common Representative: JANSSEN PHARMACEU-
TICA N.V.; Patent Department, Turnhoutseweg 30,
B-2340 Beerse (BE).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,
MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG,
SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ,
VN, YU, ZA, ZM, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),
Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),
European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR,
GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent
(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR,
NE, SN, TD, TG).

Declarations under Rule 4.17:

- as to applicant's entitlement to apply for and be granted a
patent (Rule 4.17(ii)) for all designations
- as to the applicant's entitlement to claim the priority of the
earlier application (Rule 4.17(iii)) for all designations
- of inventorship (Rule 4.17(iv)) for US only

Published:

- without international search report and to be republished
upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.

(54) Title: NOVEL BRAIN EXPRESSED GENE AND PROTEIN ASSOCIATED WITH BIPOLAR DISORDER

(57) Abstract: We previously identified 18q21.33-q23 as a candidate region for bipolar (BP) disorder and constructed a yeast arti-
ficial chromosome (YAC) contig map. In a next step we isolated and analysed all CAG/CTG repeats from this region and excluded
them from involvement in BP disorder. Here, in the process of identifying all CCG/CGG repeats from the region, we isolated three
potential CpG islands, one of which is located 1.5 kb upstream of a predicted exon of 3639 bp. Further analysis showed this was
part of a novel CpG-associated, brain-expressed gene, that we called NCAG1 (Novel CpG Associated Gene 1). Mutation analysis of
this positional and functional candidate identified two single nucleotide polymorphisms, none of which were shown to be associated
with the BP phenotype.



WO 02/101044 A2

NOVEL BRAIN EXPRESSED GENE AND PROTEIN ASSOCIATED WITH BIPOLAR DISORDER

5

FIELD OF THE INVENTION:

The invention is broadly concerned with the determination of genetic factors associated with psychiatric health. More particularly, the present invention is directed to a human
10 gene which is linked to a mood disorder or related disorder in affected individuals and their families. Specifically, the present invention is directed to a gene located on the eighteenth chromosome that is expressed in brain tissue and may be used as a diagnostic marker for bipolar disorder.

15

BACKGROUND OF THE INVENTION:

Pharmacogenetics background:

20 Every individual is a product of the interaction of their genes and the environment. Pharmacogenetics is the study of how genetic differences influence the variability in patients responses to drugs. Through the use of pharmacogenetics, we will soon be able to profile variations between individuals DNA to predict responses to a particular medicine. Target validation that will predict a well-tolerated and effective medicine for
25 a clinical indication in humans is a widely perceived problem; but the real challenge is target selection. A limited number of molecular target families have been identified, including receptors and enzymes, for which high throughput screening is currently possible. A good target is one against which many compounds can be screened rapidly to identify active molecules (hits). These hits can be developed into optimized
30 molecules (leads), which have the properties of well-tolerated and effective medicines. Selection of targets that can be validated for a disease or clinical symptom is a major problem faced by the pharmaceutical industry. The best-validated targets are those that have already produced well-tolerated and effective medicines in humans (precedent targets). Many targets are chosen on the basis of scientific hypotheses and do not lead
35 to effective medicines because the initial hypotheses are often subsequently disproved.

Two broad strategies are being used to identify genes and express their protein products for use as high-throughput targets. These approaches of genomics and genetics share technologies but represent distinct scientific tactics and investments. Discovery genomics uses the increasing number of databases of DNA sequence information to
5 identify genes and families of genes for tractable or scrollable targets that are not known to be genetically related to disease.

The advantage of information on disease-susceptibility genes derived from patients is that, by definition, these genes are relevant to the patients' genetic contributions to the
10 disease. However, most susceptibility genes will not be tractable targets or amenable to high-throughput screening methods to identify active compounds.

The differential metabolism related to the relevant gene variants can be studied in focused functional genomic and proteomic technologies to discover mechanisms of disease development or progression.

15 Critical enzymes of receptors associated with the altered metabolism can be used as targets. Gene-to-function-to-target strategies that focus on the role of the specific susceptibility gene variants on appropriate cellular metabolism become important.

Data mining of sequences from the Human Genome Project and similar programmes with powerful bioinformatic tools has made it possible to identify gene families by
20 locating domains that possess similar sequences. Genes identified by these genomic strategies generally require some sort of functional validation or relationship to a disease process. Technologies such as differential gene expression, transgenic animal models, proteomics, in situ hybridization and immunohistochemistry are used to imply relationships between a gene and a disease.

25

The major distinction between the genomic and genetic approaches is target selection, which genetically defined genes and variant-specific targets already known to be involved in the disease process. The current vogue of discovery genomics for nonspecific, wholesale gene identification, with each gene in search of a relationship to
30 a disease, creates great opportunities for development of medicines.

It is also critical to realize that the core problem for drug development is poor target selection. The screening use of unproven technologies to imply disease-related validation, and the huge investment necessary to progress each selected gene to proof

of a concept in humans, is based on an unproven and cavalier use of the word 'validation'. Each failure is very expensive in lost time and money. For example, differential gene expression (DGE) and proteomics are screening technologies that are widely used for target validation. They detect different levels and/or patterns of gene
5 and protein expression in tissues, which may be used to imply a relationship to a disease affecting that tissue.

Mood Disorder Background:

Mood disorders or related disorders include but are not limited to the following
10 disorders as defined in the Diagnostic and statistical Manual of Mental Disorders, version 4 (DSM-IV) taxonomy DSM-IV codes in parenthesis): mood disorders (296.XX,300.4,311,301.13,295.70) , schizophrenia and related disorders (295.XX,297.1,298.8,297.3,298.9), anxiety disorders (300.XX,309.81,308.3), adjustment disorders (309.XX) and personality disorders (codes 301.XX) .

15 The present invention is particularly directed to genetic factors associated with a family of mood disorders known as Bipolar (BP) spectrum disorders. Bipolar disorder (BP) is a severe psychiatric condition that is characterized by disturbances in mood, ranging from an extreme state of elation (mania) to a severe state of dysphoria (depression). Two types of bipolar illness have been described: type I BP illness (BPI) is
20 characterized by major depressive episodes alternated with phases of mania, and type II BP illness (BPII) , characterized by major depressive episodes alternating with phases of hypomania. Relatives of BP probands have an increased risk for BP, unipolar disorder (patients only experiencing depressive episodes; UP), cyclothymia (minor depression and hypomania episodes; cy) as well as for schizoaffective disorders of the
25 manic (SAm) and depressive (SAd) type. Based on these observations BP, cY, UP and SA are classified as BP spectrum disorders.

The involvement of genetic factors in the etiology of BP spectrum disorders was suggested by family, twin and adoption studies (Tsuang and Faraone (1990), the Genetics of Mood Disorders, Baltimore, The John Hopkins University Press) However,
30 the exact pattern of transmission is unknown. In some studies, complex segregation analysis supports the existence of a single major locus for BP (Spence et al. (1995), Am J.Med. Genet (Neuropsych. Genet.) QQ pp 370-376). Other researchers propose a liability-threshold-model, in which the liability to develop the disorder results from the

additive combination of multiple genetic and environmental effects (McGuffin et al. (1994) , Affective Disorders; Seminars in Psychiatric Genetics Gaskell, London pp 110-127) .

- Due to the complex mode of inheritance, parametric and non-parametric linkage strategies are applied in families in which BP disorder appears to be transmitted in a Mendelian fashion. Early linkage findings on chromosomes 11p15 (Egeland et al. (1987) , Nature ~ pp 783-787) and Xq27-q28 (Mendlewicz 'et al. (1987, the Lancet l pp 1230 -1232; Baron et al. (1987) Nature 12& pp 289-292) have been controversial and could initially not be replicated (Kelsoe et al. (1989) Nature ~ pp 238-243; Baron et al. (1993) Nature Genet ~ pp 49-55) .with the development of a human genetic map saturated with highly polymorphic markers and the continuous development of data analysis techniques, numerous new linkage searches were started. In several studies, evidence or suggestive evidence for linkage to particular regions on chromosomes 4, 12, 18, 21 and X was found (Black wood et al. (1996) Nature Genetics ~ pp 427-430, Craddock et al. (1994) Brit J. psychiatry ~ pp355-358, Berrettini et al. (1994), Proc Natl Acad Sci USA ~ pp 5918-5921, Straub et al. (1994) Nature Genetics ~ pp 291-296 and Pekkarinen et al. (1995) Genome Research 2 pp 105-115). In order to test the validity of the reported linkage results, these findings have to be replicated in other, independent studies.
- Recently, linkage of bipolar disorder to the pericentromeric region on chromosome 18 was reported (Berrettini et al. 1994). Also a ring chromosome 18 with break-points and deleted regions at 18pter-p11 and 18q23-qter was reported in three unrelated patients with BP illness or relates syndromes (Craddock et al. 1994). The chromosome 18p linkage was replicated by stine et al. (1995) Am J. Hum Genet 22 pp 1384-1394, who also reported suggestive evidence for a locus on 18q21.2-q21.32 in the same study.

- Interestingly, Stine et al. observed a parent-of-origin effect: the evidence of linkage was the strongest in the paternal pedigrees, in which the proband's father or one of the proband's father's sibs is affected. Several studies described anticipation in families transmitting BP disorder(McInnis et al 1993, Nylander et al 1994) suggesting the involvement of trinucleotide repeat expansions (TREs), considering a number of diseases caused by an expansion of a CAG/CTG, a CCG/CGG or a GAA/TTC repeat show anticipation (reviewed by Margolis et al.(Margolis et al 1999)). Previous efforts

to find potentially expanded repeats have primarily focused on CAG/CTG repeats although the search for CCG/CGG repeats is increasing(Kleiderlein et al 1998, Mangel et al 1998, Eichhammer et al 1998, Kaushik et al 2000). Previously, we reported on a new method for the region specific isolation of triplet repeats: triplet repeat YAC
5 fragmentation(Del Favero et al 1999). This proved to be a valid method for the isolation of CAG/CTG repeats and using this method, we excluded the involvement of CAG/CTG repeats from within 18q21.33-q23 in bipolar disorder(Goossens et al 2000). The present invention adapted the method for the region specific isolation of CCG/CGG repeats and applied it to the chromosome 18q21.33-q23 BP candidate
10 region.

SUMMARY OF THE INVENTION:

The present invention is directed to a novel gene and protein encoded by that gene.

The novel gene is located at an 8.9 cM chromosome region located between D18S68
15 and D18S979 at 18q21.33-q23. A physical map was constructed using yeast artificial chromosomes (YACs)(Verheyen et al 1999).

The previously described method was adapted for the region specific isolation of CCG/CGG repeats and applied to the chromosome 18q21.33-q23 BP candidate region. Three potential CpG islands were isolated, one of which is located 1.5 kb upstream of
20 a predicted exon of 3639 bp. Further analysis showed this was part of a novel CpG-associated, brain-expressed gene, herein called NCAG1 (Novel CpG Associated Gene 1). Mutation analysis of this positional and functional candidate identified two single nucleotide polymorphisms, which may be useful as a diagnostic marker for BP phenotype.

25

BRIEF DESCRIPTION OF THE DRAWING

Figure 1. List of all human ESTs found by BLASTN alignment searches of dbEST. ESTs are named with their Genbank Acc Nos. I.M.A.G.E. Consortium [LLNL] cDNA
30 Clones(Lennon et al 1996) are named with their RZPD clone ID.

Figure 2: Minimal YAC tiling path of the 18q21.33-q23 BP candidate region(Verheyen et al 1999). The YACs are represented by solid lines, the CCG/CGG

fragmentation products by dotted lines. YAC sizes, between brackets, are estimated by PFGE analysis. Solid circles indicate positive STS/STR hits. Shaded boxes highlight the CCG/CGG repeat and the three CpG islands isolated by YAC fragmentation.

- 5 **Figure 3:** Feature map of NCAG1. a) Predicted Features by bioinformatics. They encompass the CpG island as predicted by LCP(Huang 1994) and CPG(Larsen et al 1992), the ORF or exon as predicted by Grail(Uberbacher & Mural 1991) and Genscan(Burge & Karlin 1997), the transcription start site (TSS) as predicted by Proscan(Prestridge 1995)and the relevant polyadenylation signals as predicted by
10 PolyAH(Salamov & Solovyev 1997). The numbers below the features indicate the scores as returned by Proscan and PolyAH. b) Alignment of EST hits. ESTs are named with their Genbank Acc Nos. c) Alignment of cDNA clones. I.M.A.G.E. Consortium [LLNL] cDNA Clones(Lennon et al 1996) are named with their RZPD clone ID. d)
15 RT-PCR products. The grey bars represent the RT-PCR product, the thin black lines represent the sequences obtained on the nested PCRs.

DETAILED DESCRIPTION OF THE INVENTION:

The present invention is directed to a novel gene located at the 18q chromosomal
20 candidate region of chromosome 18. More specifically, the gene is located at an 8.9 cM region located between D18S68 and D18S979 at 18q21.33-q23.

The gene is located at a chromosomal region associated with mood disorders such as bipolar spectrum disorders and may therefore be useful as a diagnostic marker for bipolar spectrum disorders. The region in question when removed from the totality of
25 the human genome may also be used to locate, isolate and sequence other genes which influences psychiatric health and mood.

Isolation and identification of Identification of novel gene:

Standard procedures well-known to one skilled in the art were applied to the identified
30 YAC clones and, where applicable, to the DNA from an individual afflicted with a mood disorder as defined herein, in the process of identifying and characterizing the relevant gene. For example, the inventors are able to make use of the previously identified apparent association between trinucleotide repeat expansions (TRE) within

the human genome and the phenomenon of anticipation in mood disorders (Lindblad et al. (1995), Neurobiology of Disease 2. pp 55-62 and O'Donovan et al. (1995), Nature Genetics 1Q pp 380-381) to screen for TRE's in the selected YAC clones in order to identify candidate genes in the region of interest on human chromosome 18. A variety of other known procedures can also be applied to the said YAC clones to identify the candidate gene as discussed below.

Accordingly, in a first aspect the present invention comprises the use of an 8.9 cM region of human chromosome 18q disposed between polymorphic markers D18S68 and D18S979 or a fragment thereof for identifying at least one human gene, including mutated and polymorphic variants thereof, which is associated with mood disorders or related disorders as defined above. As will be described below, the present inventors have identified this candidate region of chromosome 18q for such a gene, by analysis of co-segregation of bipolar disease in family MAD31 with 12 STR polymorphic markers previously located between D18S51 and D18S61 and subsequent LaD score analysis.

Particular YACs covering the candidate region which may be used in accordance with the present invention are 961.h-9, 942-c.3, 766-f-12, 731-c- 7, 907.e.1, 752-g-8 and 717-d-3, preferred ones being 961h-9, 766.f.12 and 907-e.1 since these have the minimum tiling path across the candidate region. suitable YAC clones for use are those having an artificial chromosome spanning the refined candidate region between D18S68 and D18S979.

There are a number of methods which can be applied to the candidate regions of chromosome 18q as defined above, whether or not present in a YAC, to identify a candidate gene or genes associated with mood disorders or related disorders. For example, as aforesaid, there is an apparent association between the extent of trinucleotide repeat expansions (TRE) in the human genome and the presence of mood disorders.

Accordingly, in a third aspect the present invention comprises a method of identifying at least one human gene, including mutated and polymorphic variants thereof, which is associated with a mood disorder or related disorder as defined herein which comprises detecting nucleotide triplet repeats in the region of human chromosome 18q disposed between polymorphic markers D18S68 and D18S979.

- An alternative method of identifying said gene or genes comprises fragmenting a YAC clone comprising a portion of human chromosome 18q disposed between polymorphic markers D18S60 and D18S61, for example one or more of the seven aforementioned YAC clones, and detecting any nucleotide triplet repeats in said fragments, in particular repeats of CAG or CTG. Nucleic acid probes comprising at least 5 and preferably at least 10 CTG and/or CAG triplet repeats are a suitable means of detection when appropriately labelled. Trinucleotide repeats may also be determined using the known RED (repeat expansion detection) system (Shalling et al. (1993), Nature Genetics ~ pp 135-139).
- 10 In a fourth embodiment the invention comprises a method of identifying at least one gene, including mutated and polymorphic variants thereof, which is associated with a mood disorder or related disorder and which is present in a YAC clone spanning the region of human chromosome 18q between polymorphic markers D18S60 and D18S61, the method comprising the step of detecting the expression
- 15 product of a gene incorporating nucleotide triplet repeats by use of an antibody capable of recognizing a protein with an amino acid sequence comprising a string of at least 8, but preferably at least 12, continuous glutamine residues. Such a method may be implemented by sub-cloning YAC DNA, for example from the seven aforementioned YAC clones, into a human DNA expression library. A preferred means of detecting the
- 20 relevant expression product is by use of a monoclonal antibody, in particular mAB1C2, the preparation and properties of which are described in International Patent. Application Publication No WO 97/17445.
- Further embodiments of the present invention relate to methods of identifying the relevant gene or genes which involve the sub-cloning of YAC DNA as defined above
- 25 into vectors such as BAC (bacterial artificial chromosome) or PAC (P1 or phage artificial chromosome) or cosmid vectors such as exon-trap cosmid vectors. The starting point for such methods is the construction of a contig map of the region of human chromosome 18q between polymorphic markers D18S60 and D18S61. To this end the present inventors have sequenced the end regions of the fragment of human
- 30 DNA in each of the seven aforementioned YAC clones and these sequences are disclosed herein. Following sub-cloning of YAC DNA into other vectors as described above, probes comprising these end sequences or portions thereof, in particular those sequences shown in Figures 1 to 11 herein, together with any known sequenced tagged

site (STS) in this region, as described in the YAC clone contig shown herein, as can be used to detect overlaps between said sub-clones and a contig map can be constructed. Also the known sequences in the current YAC contig can be used for the generation of contig map sub-clones.

- 5 One route by which a gene or genes which is associated with a mood disorder or associated disorder can be identified is by use of the known technique of exon trapping. This is an artificial RNA splicing assay, most often making use in current protocols of a specialized exon-trap cosmid vector. The vector contains an artificial mini-gene consisting of a segment of the SV40 genome containing an origin of replication and a
10 powerful promoter sequence, two splicing-competent exons separated by an intron which contains a multiple cloning site and an SV40 polyadenylation site.

- The YAC DNA is sub-cloned in the exon-trap vector and the recombinant DNA is transfected into a strain of mammalian cells. Transcription from the SV40 promoter results in an RNA transcript which normally splices to include the two exons of the
15 minigene. If the cloned DNA itself contains a functional exon, it can be spliced to the exons present in the vector's minigene. Using reverse transcriptase a cDNA copy can be made and using specific PCR primers, splicing events involving exons of the insert DNA can be identified. Such a procedure can identify coding regions in the YAC DNA which can be compared to the equivalent regions of DNA from a person afflicted with
20 a mood disorder or related disorder to identify the relevant gene.

Accordingly, in a fifth aspect the invention comprises a method of identifying at least one human gene, including mutated variants and polymorphisms thereof, which is associated with a mood disorder or related disorder which comprises the steps of:

- (1) transfecting mammalian cells with exon trap cosmid vectors prepared and mapped
25 as described above;
(2) culturing said mammalian cells in an appropriate medium;
(3) isolating RNA transcripts expressed from the SV40 promoter;
(4) preparing cDNA from said RNA transcripts;
(5) identifying splicing events involving exons of the DNA sub-cloned into said exon
30 trap cosmid vectors to elucidate positions of coding regions in said sub-cloned DNA;
(6) detecting differences between said coding regions and equivalent regions in the DNA of an individual afflicted with said mood disorder or related disorder; and

(7) identifying said gene or mutated or polymorphic variant thereof which is associated with said mood disorder or related disorders.

As an alternative to exon trapping the YAC DNA may be sub-cloned into BAC, PAC, cosmid or other vectors and a contig map constructed as described above. There are a variety of known methods available by which the position of relevant genes on the sub-cloned DNA can be established as follows:

- (a) cDNA selection or capture (also called direct selection and cDNA selection) : this method involves the forming of genomic DNA/cDNA heteroduplexes by hybridizing a cloned DNA (e.g. an insert of a YAC DNA), to a complex mixture of cDNAs, such as the inserts of all cDNA clones from a specific (e.g. brain) cDNA library. Related sequences will hybridize and can be enriched in subsequent steps using biotin-streptavidine capturing and PCR (or related techniques);
- (b) hybridization to mRNA/cDNA: a genomic clone (e.g. the insert of a specific cosmid) can be hybridized to a Northern blot of mRNA from a panel of culture cell lines or against appropriate (e.g. brain) cDNA libraries. A positive signal can indicate the presence of a gene within the cloned fragment;
- (c) CpG island identification: CpG or HTF islands are short (about 1 kb) hypomethylated GC-rich (> 60%) sequences which are often found at the 5' ends of genes. CpG islands often have restriction sites for several rare-cutter restriction enzymes. Clustering of rare-cutter restriction sites is indicative of a CpG island and therefore of a possible gene. CpG islands can be detected by hybridization of a DNA clone to Southern blots of genomic DNA digested with rare-cutting enzymes, or by island-rescue PCR (isolation of CpG islands from YACs by amplifying sequences between islands and neighbouring Alu-repeats) ;
- (d) zoo-blotting: hybridizing a DNA clone (e.g. the insert of a specific cosmid) at reduced stringency against a Southern blot of genomic DNA samples from a variety of animal species. Detection of hybridization signals can suggest conserved sequences, indicating a possible gene. Accordingly, in a sixth aspect the invention comprises a method of identifying at least one human gene including mutated and polymorphic variants thereof which is associated with a mood disorder or related disorder which comprises the steps of:
 - (1) sub-cloning the YAC DNA as described above into a cosmid, BAC, PAC or other vector;

(2) using the nucleotide sequences shown in any one of Figures 1 to 11 or any other sequenced tagged site (STS) in this region as in the YAC clone contig described herein, or part thereof consisting of not less than 14 contiguous bases or the complement thereof, to detect overlaps amongst the sub-clones and construct a map thereof;

5 (3) identifying the position of genes within the sub-cloned DNA by one or more of CpG island identification, zoo-blotting, hybridization of the sub-cloned DNA to a cDNA library or a Northern blot of mRNA from a panel of culture cell lines;

(4) detecting differences between said genes and equivalent region of the DNA of an individual afflicted with a mood disorder or related disorder; and

10 (5) identifying said gene which is associated with said mood disorders or related disorders.

If the cloned YAC DNA is sequenced, computer analysis can be used to establish the presence of relevant genes. Techniques such as homology searching and exon prediction may be applied.

15 Once a candidate gene has been isolated in accordance with the methods of the invention more detailed comparisons may be made between the gene from a normal individual and one afflicted with a mood disorder such as a bipolar spectrum disorder. For example, there are two methods, described as "mutation testing", by which a mutation or polymorphism in a DNA sequence can be identified. In the first the DNA
20 sample may be tested for the presence or absence of one specific mutation but this requires knowledge of what the mutation might be. In the second a sample of DNA is screened for any deviation from a standard (normal) DNA. This latter method is more useful for identifying candidate genes where a mutation is not identified in advance. In addition the following techniques may be further applied to a gene identified by the
25 above-described methods to identify differences between genes from normal or healthy individuals and those afflicted with a mood disorder or related disorder:

(a) Southern blotting techniques: a clone is hybridized to nylon membranes containing genomic DNA digested with different restriction enzymes of patients and healthy individuals. Large differences between patients and healthy individuals can be
30 visualized using a radioactive labelling protocol;

(b) heteroduplex mobility in polyacrylamide gels: this technique is based on the fact that the mobility of heteroduplexes in non-denaturing polyacrylamide gels is less than the mobility of homoduplexes. It is most effective for fragments under 200 bp;

(c) single-strand conformational polymorphism analysis (SSCP or SSCA) : single stranded DNA folds up to form complex structures that are stabilized by weak intramolecular bonds.

The electrophoretic mobilities of these structures on non-denaturing polyacrylamide
5 gels depends on their chain lengths and on their conformation;

(d) chemical cleavage of mismatches (CCM) : a radiolabelled probe is hybridized to the test DNA, and mismatches detected by a series of chemical reactions that cleave one strand of the DNA at the site of the mismatch. This is a very sensitive method and can be applied to kilobase-length samples;

10 (e) enzymatic cleavage of mismatches: the assay is similar to CCM, but the cleavage is performed by certain bacteriophage enzymes.

(f) denaturing gradient gel electrophoresis: in this technique, DNA duplexes are forced to migrate through an electrophoretic gel in which there is a gradient of increasing amounts of a denaturant (chemical or temperature). Migration continues until the DNA
15 duplexes reach a position on the gel wherein the strands melt and separate, after which the denatured DNA does not migrate much further. A single base pair difference between a normal and a mutant DNA duplex is sufficient to cause them to migrate to different positions in the gel;

(g) direct DNA sequencing.

20 It will be appreciated that with respect to the methods described herein, in the step of detecting differences between coding regions from the YAC and the DNA of an individual afflicted with a mood disorder or related disorder, the said individual may be anybody with the disorder and not necessary a member of family MAD31.

25 In accordance with further aspects the present invention provides an isolated human gene and variants thereof associated with a mood disorder or related disorder and which is obtainable by any of the above described methods, an isolated human protein encoded by said gene and a cDNA encoding said protein.

30 Once a gene has been identified a number of methods are available to determine the function of the encoded protein. These methods are described by Eisenberg et al (Nature vol. 15, June 2000) and is herein incorporated by reference. One method involves a computational method that reveals functional linkages from genome

sequences and is called the gene neighbor metho. If in several genomes the genes that encode two proteins are neighbors on the chromosome, the proteins tend to be functionally linked. This method can be powerful in uncovering functional linkages in prokaryotes, where operons are common, but also shows promise for analysing
5 interacting proteins in eukaryotes.

Examples:

Example 1

10

A :Triplet repeat isolation

CCG/CGG YAC fragmentation vectors were constructed by cloning blunted (CCG)₁₀/(CGG)₁₀ adapters into the blunted SphI site of the previously described pDV1 basic vector(Del-Favero et al 1999). Sequencing determined that fragmentation vectors
15 pDVCCG and pDVCGG have the adapter sequence in a 5'-(CCG)₁₀-3' and a 5'-(CGG)₁₀-3' orientation respectively.

Using these vectors, CCG/CGG repeats and flanking sequences were isolated by YAC fragmentation as described(Del-Favero et al 1999).

20 B: Characterisation of Structure of the NCAG1 gene.

I.M.A.G.E. Consortium [LLNL] cDNA Clones(Lennon et al 1996)
IMAGp998A136826Q2, IMAGp998A154307Q2, IMAGp998B194346Q2,
IMAGp998D126826Q2, IMAGp998D193628Q2, IMAGp998F131866Q2,
IMAGp998H201815Q2, IMAGp998K235214Q2, IMAGp998L153967Q2 and
25 IMAGp998N06839Q2 were ordered at RZPD Deutsches Ressourcenzentrum für Genomforschung GmbH (Heubnerweg 6, 14059 Berlin-Charlottenburg, Germany). Cultures starting from single colonies were grown and plasmids were prepared by the Wizard Plus SV Minipreps DNA Purification System (Promega, Madison, WI). DNA sequencing was performed with the dideoxynucleotide sequencing method using a
30 DNA sequencing kit (Perkin-Elmer, Foster, CA) and analysed by an ABI PRISM 377 DNA Sequencer (Perkin-Elmer, Foster, CA) or an ABI PRISM 3700 DNA Analyser (Perkin-Elmer, Foster, CA).

For the RT-PCR reactions, mRNA from *SHSY-5Y* cells was prepared using the μ MACS mRNA Isolation Kit (Miltenyi Biotec, Bergisch Gladbach, Germany). After
35 DNaseI treatment (Promega, Madison, WI), the RT reaction was primed with

oligo(dT) primers and performed with Superscript Preamplification System for First Strand cDNA synthesis (GibcoBRL, N.V. Life Technologies, Merelbeke, Belgium). FscDNA was used in long-range PCR reactions with TaKaRa LA Taq (Takara Shuzo Co., Otsu, Shiga, Japan). PCR products were reamplified with nested primers and
5 sequenced as described above.

C: Characterisation of the expression pattern of the NCAG1 gene.

Genepool cDNA (Invitrogen, Carlsbad, CA) from brain, fetal brain, placenta, liver, testis and lung was used as a cDNA mapping panel. The Human Brain Multiple Tissue
10 Northern (MTN) Blot IV (Clontech, Palo Alto, CA) was used for radioactive hybridisation in accompanying ExpressHyb solution according to the instructions of the manufacturer. A zooblot was prepared by digesting 10 µg genomic DNA to completion with HindIII, running it on a TAE 1% agarose gel and performing a Southern blot. A PCR product containing the ORF of the NCAG1 gene was radioactively labelled and
15 hybridised at 65 °C.

D: Mutation analysis of the NCAG1 gene.

Overlapping PCR products of approximately 600 bp were generated and sequenced as described above. Both identified polymorphisms were detected by digesting the PCR
20 product with HinfI and electrophoresing the fragments on precast ExcelGel gels on a Multiphor II electrophoresis system (Amersham Pharmacia Biotech AB, Uppsala, Sweden)

E: CCG/CGG YAC fragmentation

CCG/CGG YAC fragmentation was applied to YACs 961h9, 766f12 and 907e1(Goossens et al 2000). Size determination by Pulsed Field Gel Electrophoresis (PFGE) and Southern blot hybridisation resulted in 33 sets of equally sized fragmented YAC clones. Sequencing of 112 fragmented YAC ends identified seven (out of 33) sets
30 of fragmented YACs with identical end sequences resulting from a specific homologous recombination. One set (CCG7) was the result of fragmentation in the (CGG)₆ repeat in the 5' UTR of the CAP2 gene (GenBank acc. No L40377). A second set (CCG6) contained a (CCG)₂ repeat and a third (CCG4) an imperfect CCCCCG repeat. The triplet repeat in the 5' UTR of the CAP2 gene was already shown not to be
35 associated with BP disorder(Goossens et al 2000). The size of CCG4 was analyzed in

12 BP and 12 UP patients, but only one allele was detected. The size of CCG6 was not analyzed since it was too small to be polymorphic.

In depth analysis showed that three (CCG3, GenBank acc No ...; CCG4, GenBank acc No... and CCG6, GenBank acc No ...) of the seven sequences had high CG content
5 (70-80 %) and high CpG content (15-20 CpGs in 200 bp) but no additional CCG/CGG repeats were found. Primer pairs for these potential CpG islands were used to determine their position on the YAC contig (Figure1). BLASTN analysis(Altschul et al 1990) resulted for both CCG4 and CCG6 in hits with sequences of RPCI-11 BACs. CCG4 gave a hit in a contig of 27150 bp of the working draft sequence of RPCI-11
10 BAC 29O13 (GenBank acc No AC022662, GI: 7249117). CCG6 was part of the complete sequence of RPCI-11 BAC 793J2 (GenBank acc No AC009802).

F: Identification and *in silico* characterisation of NCAG1 gene.

To find genes possibly associated with the potential CpG islands CCG4 and CCG6,
15 their surrounding BAC sequences were analysed using bioinformatic tools. Hence the 27150 bp contig of BAC 29O13 and the complete sequence of BAC 793J2 were sent for analysis to the Rummage High-Throughput Sequence Annotation Server (<http://gen100.imb-jena.de/rummage/index.html>).

First, LCP(Huang 1994) and CPG(Larsen et al 1992) recognized CpG islands
20 containing CCG4 and CCG6 of 1.2 kb and 0.4 kb respectively, confirming their potential role as CpG islands.

In a next step, exon prediction programs Grail(Uberbacher & Mural 1991) and Genscan(Burge & Karlin 1997) both predicted the presence of a 3639 bp exon, 1.5 kb downstream of the 1.2 kb large CpG island containing CCG4. This predicted exon
25 contains an open reading frame (ORF) which starts at an ATG start codon with an almost perfect Kozak sequence and ends with a TAA stop codon. Other predicted features are a transcription start site (TSS) at 2352 bp upstream of the ORF (score 76.6 by Proscan(Prestridge 1995)) and polyadenylation signals at 3032, 3247, 4364, 5338 and 8266 downstream of the ORF (respective scores of 4.79, 3.83, 4.94, 4.93 and 6.27
30 by PolyAH(Salamov & Solovyev 1997)) (Figure2a).

BLASTN(Altschul et al 1990) alignment searches to sequences of dbEST revealed significant homology ($\geq 97\%$) to 21 human ESTs (Table1, Figure2b). TBLASTX(Altschul et al 1997) searches of the Genbank non-redundant database (nr)

with the ORF showed extensive homology on protein level with SART-2 (Genbank Acc No NP_037484), a squamous cell carcinoma antigen recognized by T-cells(Nakao et al 2000). Weaker homology was found with a series of sulfotransferases. Analysis of the 1212 long aminoacid sequence of the translated ORF by SMART (Simple Modular
5 Architecture Research Tool, V3.1)(Schultz et al 2000) did not result in any known domains apart from a cleavable signal peptide at position 1-20 and two transmembrane segments at positions 771-791 and 800-820. Interpro reported no significant hits, although BLASTP(Altschul et al 1997) of the Prodom database showed homology between the NCAG1 gene and the chondroitin-6-sulfotransferase domain (Prodom Acc
10 No PD042460)

G: Characterisation of the structural organisation of the NCAG1 gene.

Based on the BLASTN EST hits I.M.A.G.E. Consortium [LLNL] cDNA Clones(Lennon et al 1996) were ordered and sequenced. The sequences alligned with
15 the genomic sequence in the presumed 5' UTR (untranslated region), the ORF and the presumed 3' UTR, indicating that these sequences are indeed transcribed (Figure2c). Alignment of the sequence of IMAGp998B194346Q2 with the genomic sequence showed that a 865 bp fragment was missing in the cDNA. A detailed analysis of the flanking sequences revealed the presence of consensus acceptor and donor splice sites,
20 confirming that this fragment is probably an intron. Also clone IMAGp998D193628Q2 missed a fragment of 1.9 kb when compared to the genomic sequence, but consensus splice sites were absent. Two clones, IMAGp998D193628Q2 and IMAGp998A136826Q2, terminated exactly at the predicted polyadenylation signal, 4.4 kb downstream of the ORF. Sequences of clones IMAGp998A154307Q2,
25 IMAGp998D126826Q2 and IMAGp998F131866Q2 did not align with the genomic sequence and were not analysed further.

Since cDNA clone sequencing did not result in a continuous sequence of the transcript, primers were designed and used for RT-PCR experiments. Sequencing of different overlapping RT-PCR products confirmed the presence of a transcript of at least 9 kb,
30 containing the ORF of the predicted exon, linked to the presumed 5' and 3' sequences (Figure2d). The 5 prime intron of 865 bp was confirmed and the 3' UTR was extended till the predicted polyadenylation signal, 4.4 kb downstream of the ORF.

H: Characterisation of the expression pattern of the NCAG1 gene.

To investigate the expression profile of the NCAG1 gene, a long-range PCR spanning the ORF was optimised on genomic DNA and applied on a cDNA mapping panel. This showed that the fragment was present in cDNA from brain, fetal brain, placenta and
5 liver but could not be detected in cDNA from testis and lung. More detailed information on the expression in the brain was obtained by Northern blot hybridisation showing expression of a ≥ 9.5 kb transcript in all investigated tissues (lung, placenta, small intestine, liver, kidney, skeletal muscle, heart, brain, uterus, trachea, thyroid, stomach, spinal cord, prostate, mammary gland, lymph node, brain (whole), bladder,
10 adrenal gland, amygdala, caudate nucleus, corpus callosum, hippocampus, substantia nigra, thalamus and total brain).

Stringent Zooblot hybridisation experiments showed the presence of homologous sequences in the genomic DNA of other mammals like dog, pig, mouse, donkey, horse and sheep.

15

I: Mutation analysis of the NCAG1 gene.

Since this novel CpG-associated gene is brain-expressed and located in the chromosome 18q21.3-q23 BP candidate region, a mutation analysis of the ORF was performed on 3 patients and 1 escapee of the chromosome 18 linked family MAD31. In
20 this way two single nucleotide polymorphisms were identified. The first is a C to T transition on position 2017 of the ORF, changing aminoacid (AA) 673 from proline to serine. This polymorphism was only found in the healthy control. The second polymorphism was found in all three patients. It was also a C to T transition, located at position 2824 and changing the 942 AA from proline to serine. Analysis of this
25 polymorphism in family MAD31 showed that the T-allele was present on the disease haplotype.

Both polymorphisms were analysed in an association study on 92 BP patients and 92 age, sex and ethnicity matched controls by PCR-RFLP analysis. The P673S polymorphism turned out to be a frequent polymorphism with both alleles roughly
30 equally present. The P942S polymorphism however was found to be a rare polymorphism, with the T allele only present in 3 BP patients and in 2 controls. Statistical analysis showed the control population was in Hardy-Weinberg equilibrium for both polymorphisms. No alleles, genotypes or haplotypes were found to be associated to BP disorder.

Since triplet repeat fragmentation was proven to be a valid method for the region specific isolation of triplet repeats(Goossens et al 2000), we applied it to the chromosome 18q21.33-q23 BP candidate region for the isolation of CCG/CGG repeats.

5 Therefore, we first had to construct a new set of fragmentation vectors, pDVCCG and pDVCGG. Fragmentation experiments with these vectors resulted in transformation and fragmentation efficiencies in the same range as obtained with the CAG/CTG fragmentation vectors pDVCAG and pDVCTG (data not shown). Application of CCG/CGG fragmentation to YAC 961h9 resulted in the isolation of the (CGG)₆ repeat
10 in the 5' UTR of *CAP2*. This repeat is adjacent to the (CAG)₆ repeat previously reported(Goossens et al 2000). There, it was shown that this (CGG)₆(CAG)₆ repeat is polymorphic but not expanded in BP cases nor associated with BP disorder. Taken together, the CCG/CGG YAC fragmentation data does not support CCG/CGG repeats as disease causing agents in chromosome 18q21.33-q23 linked BP disorder.

15 On the other hand, fragmentation experiments resulted in three sequences (CCG3, CCG4 and CCG6) with high CG (70 – 80 %) and CpG content but containing no CCG/CGG repeat. CpG islands are usually defined as regions of DNA of more than 200 bases that have a CG content above 50 % and a ratio of observed versus expected CpGs close to that statistically expected. Therefore, CCG3, CCG4 and CCG6 can be
20 considered as potential CpG islands. Analysis of surrounding sequences of CCG4 and CCG6 with LCP(Huang 1994) and CPG(Larsen et al 1992) confirmed that the fragmentation occurred in both cases indeed in a CpG island. Since CpG islands are strongly associated with genes, more specifically housekeeping and widely expressed genes, these three sequences are likely to be located near this class of genes.

25 In the search for genes possibly associated with the isolated CpG islands, exon prediction programs Grail(Uberbacher & Mural 1991) and Genscan(Burge & Karlin 1997) both predicted the presence of a 3.6 kb exon downstream of the largest CpG island isolated. Two facts argued strongly against a false positive prediction. The first was that this two programs, based on different models, predicted exactly the same
30 exon. The second was the mere presence in genomic DNA of this ORF continuing for 3.6 kb and starting with a Kozak consensus ATG. Additional evidence that this exon was indeed transcribed was found in the fact that a series of ESTs had very high homologies (97-100 %) with sequences in and surrounding the ORF. In a next step, this

- evidence was extended by sequencing of the cDNA clones from which the ESTs originated. The EST sequences were prolonged and corrected and the homologies increased to 99-100 %. The fact that the cDNA clones originated from different cDNA libraries (Table1) indicated that the gene was expressed in different tissues. RT-PCR and northern blot experiments resulted in the final confirmation that this ORF was widely expressed, a usual characteristic of a CpG-associated gene.
- cDNA clone sequencing resulted in complete sequence of seven human cDNA clones aligning with NCAG1. In two cases a piece of genomic DNA was missing in the cDNA sequence. Clone IMAGp998B194346Q2 lacked a 865 bp fragment (Figure2c). Since this fragment was flanked by splice donor and acceptor consensus sequences, and since the fragment was also missing in the RT-PCR products, enough evidence was gathered to call it an intron. Clone IMAGp998D193628Q2 also missed a 1.4 kb fragment compared to the genomic sequence. In this case no consensus splice sites were present. Moreover cDNA clones IMAGp998L153967Q2 and IMAGp998A136826Q2 contain sequences that are located in the missing fragment of IMAGp998D193628Q2 (Figure2c). This data together with the fact that EST AA442543 is located entirely in the missing fragment (Figure2b) and the presence of this fragment in the RT-PCR products (Figure2d) indicate that this fragment might rather be an artifact than an intron.
- EST-homologies and cDNA clone sequencing proved that a series of cDNA clones terminated at a predicted polyadenylation signal, 4.3 kb downstream of the ORF or 10.3 kb downstream of the predicted TSS. If the 5 prime intron of 865 bp is taken into account, the size of transcript will be 9.5 kb, which is the size of the transcript recognized in the Northern blot experiment.
- On protein level, a cleavable signal peptide and two transmembrane domains are predicted. If this is correct, both N-terminal and C-terminal sides will be at the same side of the membrane in which it is embedded. The strong homology with the SART-2 protein is significant, but it does not add more clues as to potential functions of the novel protein.
- The 2824T allele, present on the disease haplotype in the chromosome 18 linked family MAD31, is a very rare allele with a frequency of 0.03. Therefore statistical analysis in an association sample loses a lot of its strength, leaving the possibility that this allele confers an increased risk for BP disorder.

REFERENCES

The following references are herein expressly incorporated by reference:

- 5 1. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403-10
2. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein
10 database search programs. *Nucleic Acids Res.* 25(17):3389-402
3. Burge C, Karlin S. 1997. Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* 268(1):78-94
4. Del-Favero J, Goossens D, Van den Bossche D, Van Broeckhoven C. 1999. YAC
15 fragmentation with repetitive and single-copy sequences: detailed physical mapping of the presenilin 1 gene on chromosome 14. *Gene* 229:193-201
5. Del Favero J, Goossens D, De Jonghe P, Benson K, Michalik A, Van den BD, Horwitz M, Van Broeckhoven C. 1999. Isolation of CAG/CTG repeats from within the chromosome 2p21-p24 locus for autosomal dominant spastic paraplegia (SPG4) by YAC fragmentation. *Hum. Genet.*
20 105(3):217-25
6. Eichhammer P, Walz A, Mengling T, Scholer A, Putzhammer A, Rohrmeier T, Aigner JM, Klein HE, Schlegel J. 1998. Detection of polymorphic triplet repeats in the genomes of patients suffering from bipolar affective disorder. *Int. J. Mol. Med.* 1(6):989-93
- 25 7. Goossens D, Villafuerte S, Tissir F, Van Gestel S, Claes S, Souery D, Massat I, Van den Bossche D, Van Zand K, Mendlewicz J, Van Broeckhoven C, Del-Favero J. 2000. No evidence for the involvement of CAG/CTG repeats from within 18q21.33-q23 in bipolar disorder. *Eur. J. Hum. Genet.* 8(5):385-8
- 30 8. Huang X. 1994. An algorithm for identifying regions of a DNA sequence that satisfy a content requirement. *Comput. Appl. Biosci.* 10(3):219-25

9. Kaushik N, Malaspina A, de Belleruche J. 2000. Characterization of trinucleotide- and tandem repeat-containing transcripts obtained from human spinal cord cDNA library by high-density filter hybridization. *DNA Cell Biol.* 19(5):265-73
- 5 10. Kleiderlein JJ, Nisson PE, Jessee J, Li WB, Becker KG, Derby ML, Ross CA, Margolis RL. 1998. CCG repeats in cDNAs from human brain. *Hum. Genet.* 103(6):666-73
11. Larsen F, Gundersen G, Lopez R, Prydz H. 1992. CpG islands as gene markers in the human genome. *Genomics* 13(4):1095-107
- 10 12. Lennon G, Auffray C, Polymeropoulos M, Soares MB. 1996. The I.M.A.G.E. Consortium: an integrated molecular analysis of genomes and their expression. *Genomics* 33(1):151-2
13. Mangel L, Ternes T, Schmitz B, Doerfler W. 1998. New 5'-(CGG)_n-3' repeats in the human genome. *J. Biol. Chem.* 273(46):30466-71
- 15 14. Margolis RL, McInnis MG, Rosenblatt A, Ross CA. 1999. Trinucleotide repeat expansion and neuropsychiatric disease. *Arch. Gen. Psychiatry* 56(11):1019-31
15. McInnis MG, McMahon FJ, Chase GA, Simpson SG, Ross CA, DePaulo JRJ. 1993. Anticipation in bipolar affective disorder. *Am. J. Hum. Genet.* 53:385-90
- 20 16. Nakao M, Shichijo S, Imaizumi T, Inoue Y, Matsunaga K, Yamada A, Kikuchi M, Tsuda N, Ohta K, Takamori S, Yamana H, Fujita H, Itoh K. 2000. Identification of a gene coding for a new squamous cell carcinoma antigen recognized by the CTL. *J. Immunol.* 164(5):2565-74
- 25 17. Nylander PO, Engstrom C, Chotai J, Wahlstrom J, Adolfsson R. 1994. Anticipation in Swedish families with bipolar affective disorder. *J. Med. Genet.* 31:686-9
18. Prestridge DS. 1995. Predicting Pol II promoter sequences using transcription factor binding sites. *J. Mol. Biol.* 249(5):923-32

19. Salamov AA, Solovyev VV. 1997. Recognition of 3'-processing sites of human mRNA precursors. *Comput. Appl. Biosci.* 13(1):23-8
20. Schultz J, Copley RR, Doerks T, Ponting CP, Bork P. 2000. SMART: a web-based tool for the study of genetically mobile domains. *Nucleic Acids Res.* 28(1):231-4
21. Uberbacher EC, Mural RJ. 1991. Locating protein-coding regions in human DNA sequences by a multiple sensor-neural network approach. *Proc. Natl. Acad. Sci. U. S. A* 88(24):11261-5
22. Van Broeckhoven C, Verheyen G. 1999. Report of the chromosome 18 workshop. *Am. J. Med. Genet.* 88(3):263-70
23. Verheyen GR, Villafuerte SM, Del-Favero J, Souery D, Mendlewicz J, Van Broeckhoven C, Raeymaekers P. 1999. Genetic refinement and physical mapping of a chromosome 18q candidate region for bipolar disorder. *Eur. J. Hum. Genet.* 7(4):427-34

CLAIMS

What is claimed is:

5

1. An isolated nucleic acid comprising the nucleotide sequence of SEQ ID N0: 1.

2. An isolated nucleic acid consisting essentially of the nucleotide sequence of SEQ ID N0: 1.

10

3. An isolated nucleic acid for comprising a nucleotide sequence that encodes the amino acid sequence of SEQ ID N0: 2.

4. An isolated nucleic acid comprising the nucleotide sequence of SEQ ID N0: 3.

15

5. An isolated nucleic acid consisting essentially of the nucleotide sequence of SEQ ID N0: 3.

20

6. An isolated nucleic acid consisting of the nucleotide sequence of SEQ ID N0: 1 or a contiguous fragment thereof wherein said isolated nucleic acid encodes a polypeptide having biological activity of bipolar disorder protein.

25

7. An isolated nucleic acid that hybridizes under high stringency conditions to a nucleic acid having a sequence complementary to the nucleotide sequence of SEQ ID N0: 1, wherein said isolated nucleic acid encodes a polypeptide having biological activity.

30

8. An isolated nucleic acid that encodes a polypeptide having the biological activity, said isolated nucleic acid consisting of a nucleotide sequence that is at least 90% identical to the nucleotide sequence of SEQ ID N0: 1.

35

9. An isolated nucleic acid consisting of the nucleotide sequence of SEQ ID N0: 3 or a contiguous fragment thereof wherein said isolated nucleic acid encodes a polypeptide having biological activity.

10. An isolated nucleic acid that hybridizes under high stringency conditions to a nucleic acid having a sequence complementary to the nucleotide sequence of SEQ ID N0: 3, wherein said isolated nucleic acid encodes a polypeptide having the biological activity.
- 5 11. An isolated nucleic acid that encodes a polypeptide having the biological activity,, said isolated nucleic acid consisting of a nucleotide sequence that is at least 90% identical to the nucleotide sequence of SEQ ID N0: 3.
- 10 12. Isolated and substantially purified protein encoded by the nucleic acid of Claim 6.
13. Isolated and substantially purified viral inhibitory protein 1 and 2 encoded by the nucleic acid of claim 9.
- 15 14. Isolated and substantially purified viral inhibitory protein having the amino acid sequence of SEQ ID N0: 2.
15. Isolated and substantially purified protein having an amino acid sequence that is at least 90% identical to the sequence of SEQ ID N0:2.
- 20 16. Isolated and substantially purified protein having an amino acid sequence that is at least 90% identical to the sequence of SEQ ID N0:4.
17. Isolated and substantially purified protein having an amino acid sequence that is at least 90% identical to the sequence of SEQ ID N0: 4.
- 25 18. A vector comprising the nucleic acid of claim 1.
19. A vector comprising the nucleic acid of claim 4.
- 30 20. A vector comprising the nucleic acid of claim 6 operable linked to an expression control sequence.
21. A host cell comprising the nucleic acid of claim 6.
- 35 22. A host cell comprising the vector of Claim 20.

23. A method of making protein 1 and 2 comprising:

- a) introducing the nucleic acid of claim 6 into a host cell;
- b) maintaining said host cell under conditions whereby said nucleic acid is expressed to protein;
- 5 c) recovering said protein.

24. A method of making protein comprising:

- a) introducing the nucleic acid of claim 9 into a host cell;
- b) maintaining said host cell under conditions whereby said nucleic acid is expressed to
- 10 produce protein;
- c) recovering said protein.

25. A method of making protein comprising:

- a) introducing the nucleic acid of Claim 16 into a host cell;
- 15 b) maintaining said host cell under conditions whereby said nucleic acid is expressed to produce viral inhibitory protein;
- c) recovering said protein.

26. A composition comprising purified protein and a carrier.

20

27. The composition according to claim 26 which further comprises viral inhibitory protein 2.

FIG 1

EST	cDNA library	I.M.A.G.E. clone
AA022684	Soares fetal heart NbHH19W	IMAGp998N06839Q2
AA022803	Soares fetal heart NbHH19W	IMAGp998N06839Q2
AA374532	HSC172 cells I	
AA421254	Soares ovary tumor NbHOT	IMAGp998H201815Q2
AA421255	Soares ovary tumor NbHOT	IMAGp998H201815Q2
AA442543	Soares total fetus Nb2HF8 9w	IMAGp998F131866Q2
AA858162	NCI CGAP Co8	IMAGp998D193628Q2
AI088531	Soares pregnant uterus NbHPU	IMAGp998A154307Q2
AI139422	Soares pregnant uterus NbHPU	IMAGp998B194346Q2
AI168185	Soares NSF F8 9W OT PA P S1	IMAGp998L153967Q2
AI401481	Soares NhHMPu S1	IMAGp998K235214Q2
AI732945	NCI CGAP Co8	IMAGp998D193628Q2
AI791264	NCI CGAP Co8	IMAGp998D193628Q2
AW015616	NCI CGAP Sub1	
AW139834	NCI CGAP Sub3	
AW450489	NCI CGAP Sub5	
BE139460	NCI CGAP Ut4	IMAGp998A136826Q2
BE139503	NCI CGAP Ut4	IMAGp998D126826Q2
C01922	Human adult	
C03813	Human heart	
R57132	Fetal heart	

FIG 2

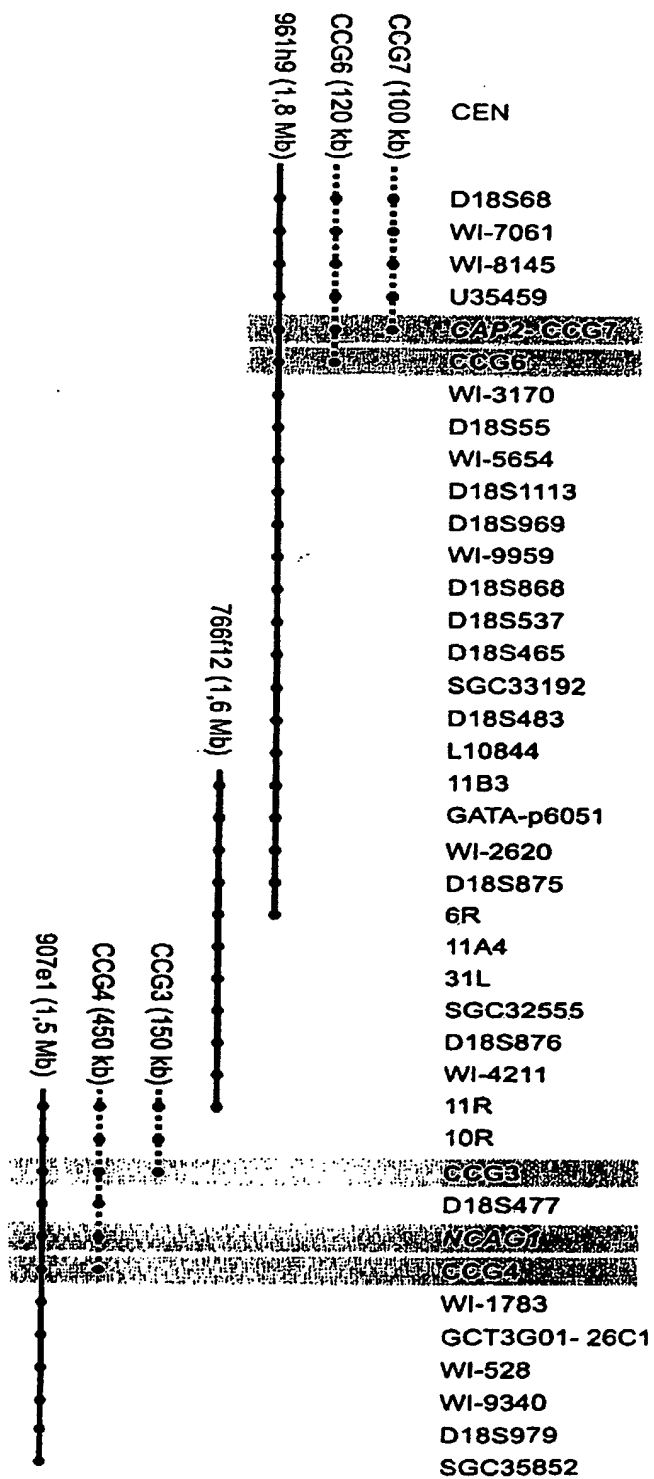
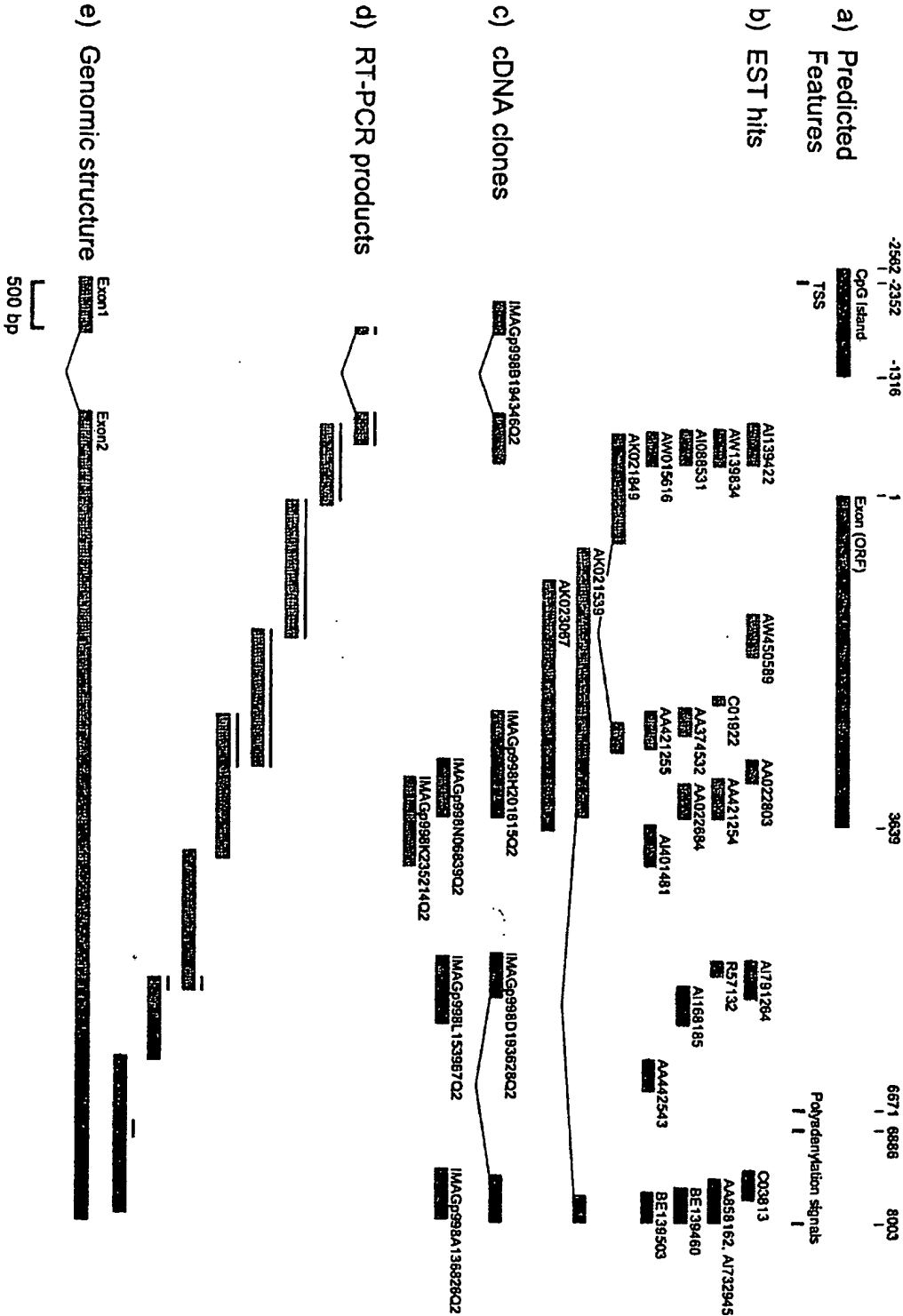


FIG 3



SEQUENCE LISTING

5 <110> Janssen Pharamceutica NV
<120> Novel Brain Expressed Gene and Protein associated with
Bipolar Disorder
10 <130> NCAG1
<140>
<141>
<160> 4
15 <170> PatentIn Ver. 2.1
<210> 1
<211> 9528
20 <212> DNA
<213> Homo sapiens
<220>
<221> CDS encoding Human NCAG1 protein
25 <222> (1507)..(5142)
<400> 1
acctgctttc ggccccgccc cgcccgccgc cggcctgctc acggctcctc cgtctctccc 60
30 cgaagccccg cctctgaccc cgccctgtcc tgtctccgtc ccgccccacg cccgccagcc 120
agcgtcgctg tctctgcct tccctgaggc ccgccttca gccccgcctt caaccccgcc 180
ccgtctgcc tccgccccgc ccccgcttgc cgccccgct cgccgtctct caccctcccc 240
35 gggctgcgcg gccggagctg gcacagagga tcctcggccg cggcgacatc accgcctggg 300
gacgcggggc ctgctctgga tacggcgcca ccgagagaac ccgccgcccg cgggtctctg 360
40 tcctgcggtc cgtggttgcc cccacaagcg tccggcggtt cctgagggcg ggcgtgtccg 420
ggccgtgcgg gtcgcgggga ccgagcgcg ctgaggagac cgagcctggg gcagcgcctg 480
ccgtagcgcg ggagacgacg cgggggtctt gcggagcccc gcgggagcct ggccccgccg 540
45 gcagagcagt tttctggaac tctccacctc cgtctccctt gggggccagt gcggcgccga 600
gcccccgctc ggatctgcct gagaaagtgt catgaaaaaa gagcagaaga gagacctcac 660
50 tgttgctgaa aggggaattt tctttcgccc gttggcggtt acttcatgat cggacgagaa 720
gtatctaggt gactgaagat attccatttt tatgtttgta cacatgaagc tgataaaaga 780
agatgtgaac atgatttctc tttgtcataa taggctgatg agtaagtaag cctgaaaaat 840
55 atttgaaatg aaggcaagaa ttttgaattt ttaaaaacca actaagactt tgatcacttg 900
ttgaggatgt ttctctctca taaatgaaag aaaaacgtat tcacaagaca agaagtataa 960
60 aaagttgaga ggaatgacaa ctgagtcac tactcgaag aatgtcagta cttcatcatc 1020
ttctttgggc aaacatacac aaatgcatca tacatgtgtg gtgagcttat caccagtgat 1080

```

ggttttctgt gctagaaatg actcttaatt tgaattttgg agtgcttttt ctcttttttt 1140
acaatgtgtg ttccaactct ttgtgttaaa tagattttaag taaaggaggt aaatgctaaa 1200
5  ttcatagtgt tttttacctg tatcacttcc ctgtgtatta tggaaaaatt agagatttta 1260
acgttattca aagttttact ggaagcaaaa ctgtgccagg gacagagata tacaatttaa 1320
gtttctcttt ttggcaactg cacttgctta naatgtactg aatgtcagct ggatttcaca 1380
10 gcatatcaga ttacagtct ttgtcttatt aaggccttta ctgtatgttt tatactaacc 1440
agatgggaaa cacattgagc atcatatctg acatgtatgc ctaagggagg agctcccca 1500
15 tggatc atg gcg tta atg ttt aca gga cat tta cta ttc tta gca tta 1548
    Met Ala Leu Met Phe Thr Gly His Leu Leu Phe Leu Ala Leu
      1          5          10
    ttg atg ttt gct ttc tct act ttt gag gaa tct gtg agc aat tat tcc 1596
    Leu Met Phe Ala Phe Ser Thr Phe Glu Glu Ser Val Ser Asn Tyr Ser
      15          20          25          30
    gaa tgg gca gtt ttc aca gat gat ata gat cag ttt aaa aca cag aaa 1644
    Glu Trp Ala Val Phe Thr Asp Asp Ile Asp Gln Phe Lys Thr Gln Lys
      35          40          45
    gtg caa gat ttc aga ccc aac caa aag ctg aag aaa agt atg ctt cat 1692
    Val Gln Asp Phe Arg Pro Asn Gln Lys Leu Lys Lys Ser Met Leu His
      50          55          60
    cca agt tta tat ttt gat gct gga gaa atc caa gca atg aga caa aag 1740
    Pro Ser Leu Tyr Phe Asp Ala Gly Glu Ile Gln Ala Met Arg Gln Lys
      65          70          75
    tct cgt gca agc cat ttg cat ctt ttt aga gct atc aga agt gca gtg 1788
    Ser Arg Ala Ser His Leu His Leu Phe Arg Ala Ile Arg Ser Ala Val
      80          85          90
    aca gtt atg ctg tcc aac cca aca tac tac cta cct cca cca aag cat 1836
    Thr Val Met Leu Ser Asn Pro Thr Tyr Tyr Leu Pro Pro Pro Lys His
      95          100          105          110
    gct gat ttt gct gcc aag tgg aat gaa att tat ggt aac aat ctg cct 1884
    Ala Asp Phe Ala Ala Lys Trp Asn Glu Ile Tyr Gly Asn Asn Leu Pro
      115          120          125
    cct tta gca ttg tac tgt ttg tta tgc cca gaa gac aaa gtt gcc ttt 1932
    Pro Leu Ala Leu Tyr Cys Leu Leu Cys Pro Glu Asp Lys Val Ala Phe
      130          135          140
    gaa ttt gtc ttg gaa tat atg gac agg atg gtt ggc tac aaa gac tgg 1980
    Glu Phe Val Leu Glu Tyr Met Asp Arg Met Val Gly Tyr Lys Asp Trp
      145          150          155
    cta gta gag aat gca cca gga gat gag gtt cca att ggc cat tcc tta 2028
    Leu Val Glu Asn Ala Pro Gly Asp Glu Val Pro Ile Gly His Ser Leu
      160          165          170
    aca ggt ttt gcc act gcc ttt gac ttt tta tat aac tta tta gat aat 2076
    Thr Gly Phe Ala Thr Ala Phe Asp Phe Leu Tyr Asn Leu Leu Asp Asn
      175          180          185          190
    cat cga aga caa aaa tac ctg gaa aaa ata tgg gtt att act gag gaa 2124

```

	His	Arg	Arg	Gln	Lys	Tyr	Leu	Glu	Lys	Ile	Trp	Val	Ile	Thr	Glu	Glu	
					195					200					205		
5	atg	tac	gag	tat	tcc	aag	gtc	cgc	tca	tgg	ggc	aaa	cag	ctt	ctc	cat	2172
	Met	Tyr	Glu	Tyr	Ser	Lys	Val	Arg	Ser	Trp	Gly	Lys	Gln	Leu	Leu	His	
				210					215					220			
10	aac	cac	caa	gcc	act	aat	atg	ata	gca	tta	ctc	aca	ggg	gcc	ttg	gtg	2220
	Asn	His	Gln	Ala	Thr	Asn	Met	Ile	Ala	Leu	Leu	Thr	Gly	Ala	Leu	Val	
			225					230					235				
15	act	gga	gta	gat	aaa	gga	tct	aaa	gca	aat	ata	tgg	aaa	cag	gct	gta	2268
	Thr	Gly	Val	Asp	Lys	Gly	Ser	Lys	Ala	Asn	Ile	Trp	Lys	Gln	Ala	Val	
		240					245					250					
20	gtg	gat	gtc	atg	gaa	aag	aca	atg	ttt	cta	ttg	aat	cat	att	gtt	gat	2316
	Val	Asp	Val	Met	Glu		Thr	Met	Phe	Leu	Leu	Asn	His	Ile	Val	Asp	
	255					260					265				270		
25	ggg	tct	ttg	gat	gaa	ggg	gtg	gcc	tat	gga	agc	tac	aca	gct	aaa	tcc	2364
	Gly	Ser	Leu	Asp	Glu	Gly	Val	Ala	Tyr	Gly	Ser	Tyr	Thr	Ala	Lys	Ser	
					275					280					285		
30	gtc	aca	cag	tat	gtt	ttt	ctg	gcc	cag	cgc	cat	ttt	aat	atc	aac	aac	2412
	Val	Thr	Gln	Tyr	Val	Phe	Leu	Ala	Gln	Arg	His	Phe	Asn	Ile	Asn	Asn	
				290					295					300			
35	ttg	gat	aat	aac	tgg	tta	aag	atg	cac	ttt	tgg	ttc	tat	tat	gcc	acc	2460
	Leu	Asp	Asn	Asn	Trp	Leu	Lys	Met	His	Phe	Trp	Phe	Tyr	Tyr	Ala	Thr	
			305					310					315				
40	ctt	tta	cct	ggc	ttc	caa	aga	act	gtg	ggg	ata	gca	gat	tcc	aat	tat	2508
	Leu	Leu	Pro	Gly	Phe	Gln	Arg	Thr	Val	Gly	Ile	Ala	Asp	Ser	Asn	Tyr	
		320					325					330					
45	aat	tgg	ttt	tat	ggg	cca	gaa	agc	cag	cta	gtt	ttc	ttg	gat	aag	ttc	2556
	Asn	Trp	Phe	Tyr	Gly	Pro	Glu	Ser	Gln	Leu	Val	Phe	Leu	Asp	Lys	Phe	
	335					340					345				350		
50	atc	tta	aag	aat	gga	gct	gga	aat	tgg	tta	gct	cag	caa	att	aga	aag	2604
	Ile	Leu	Lys	Asn	Gly	Ala	Gly	Asn	Trp	Leu	Ala	Gln	Gln	Ile	Arg	Lys	
					355				360						365		
55	cac	cga	cct	aaa	gat	gga	ccg	atg	gtt	cct	tca	act	gcc	caa	agg	tgg	2652
	His	Arg	Pro	Lys	Asp	Gly	Pro	Met	Val	Pro	Ser	Thr	Ala	Gln	Arg	Trp	
				370				375						380			
60	agt	act	ctt	cac	act	gaa	tac	atc	tgg	tat	gat	ccc	cag	ctc	aca	cca	2700
	Ser	Thr	Leu	His	Thr	Glu	Tyr	Ile	Trp	Tyr	Asp	Pro	Gln	Leu	Thr	Pro	
			385					390					395				
65	cag	cca	cct	gct	gat	tat	ggg	act	gca	aaa	ata	cac	aca	ttc	cct	aac	2748
	Gln	Pro	Pro	Ala	Asp	Tyr	Gly	Thr	Ala	Lys	Ile	His	Thr	Phe	Pro	Asn	
		400					405					410					
70	tgg	ggg	gtg	gtt	act	tat	ggg	gct	ggg	ttg	cca	aac	aca	cag	acc	aac	2796
	Trp	Gly	Val	Val	Thr	Gly	Gly	Ala	Gly	Leu	Pro	Asn	Thr	Gln	Thr	Asn	
	415					420					425				430		
75	acc	ttt	gtg	tct	ttt	aaa	tct	ggg	aag	ctg	ggg	gga	cga	gct	gtg	tat	2844
	Thr	Phe	Val	Ser	Phe	Lys	Ser	Gly	Lys	Leu	Gly	Gly	Arg	Ala	Val	Tyr	
					435					440					445		

		gac	ata	gtt	cat	ttt	cag	cca	tat	tcc	tgg	att	gat	ggg	tgg	aga	agt	2892
		Asp	Ile	Val	His	Phe	Gln	Pro	Tyr	Ser	Trp	Ile	Asp	Gly	Trp	Arg	Ser	
				450						455					460			
5		ttt	aac	cca	gga	cat	gag	cat	cca	gat	cag	aac	tca	ttt	act	ttt	gcc	2940
		Phe	Asn	Pro	Gly	His	Glu	His	Pro	Asp	Gln	Asn	Ser	Phe	Thr	Phe	Ala	
				465					470					475				
10		ccc	aat	gga	caa	gta	ttt	gtt	tct	gaa	gct	ctc	tat	gga	ccc	aag	ttg	2988
		Pro	Asn	Gly	Gln	Val	Phe	Val	Ser	Glu	Ala	Leu	Tyr	Gly	Pro	Lys	Leu	
			480					485					490					
15		agc	cac	ctt	aac	aat	gta	ttg	gtg	ttt	gct	cca	tca	ccc	tca	agc	cag	3036
		Ser	His	Leu	Asn	Asn	Val	Leu	Val	Phe	Ala	Pro	Ser	Pro	Ser	Ser	Gln	
		495					500					505					510	
20		tgt	aat	aag	ccc	tgg	gaa	ggt	caa	ctg	gga	gaa	tgt	gcg	cag	tgg	ctt	3084
		Cys	Asn	Lys	Pro	Trp	Glu	Gly	Gln	Leu	Gly	Glu	Cys	Ala	Gln	Trp	Leu	
						515					520					525		
		aag	tgg	act	ggc	gag	gag	gtt	ggt	gat	gca	gct	ggg	gaa	ata	atc	act	3132
		Lys	Trp	Thr	Gly	Glu	Glu	Val	Gly	Asp	Ala	Ala	Gly	Glu	Ile	Ile	Thr	
					530					535					540			
25		gcc	tct	caa	cat	ggg	gaa	atg	gta	ttt	gtg	agt	ggg	gaa	gcc	gtg	tct	3180
		Ala	Ser	Gln	His	Gly	Glu	Met	Val	Phe	Val	Ser	Gly	Glu	Ala	Val	Ser	
				545					550					555				
30		gct	tat	tct	tca	gca	atg	aga	ctg	aaa	agt	gta	tat	cgt	gct	ttg	ctt	3228
		Ala	Tyr	Ser	Ser	Ala	Met	Arg	Leu	Lys	Ser	Val	Tyr	Arg	Ala	Leu	Leu	
			560					565					570					
35		ctc	tta	aat	tcc	caa	act	ctg	cta	gtt	gtt	gat	cat	att	gag	agg	caa	3276
		Leu	Leu	Asn	Ser	Gln	Thr	Leu	Leu	Val	Val	Asp	His	Ile	Glu	Arg	Gln	
		575					580					585					590	
40		gaa	gat	tcc	cca	ata	aat	tct	gtc	agt	gcc	ttc	ttt	cat	aat	ttg	gat	3324
		Glu	Asp	Ser	Pro	Ile	Asn	Ser	Val	Ser	Ala	Phe	Phe	His	Asn	Leu	Asp	
						595					600					605		
		att	gat	ttt	aaa	tat	atc	cca	tat	aag	ttt	atg	aat	agg	tat	aat	ggg	3372
		Ile	Asp	Phe	Lys	Tyr	Ile	Pro	Tyr	Lys	Phe	Met	Asn	Arg	Tyr	Asn	Gly	
					610					615					620			
45		gcc	atg	atg	gat	gtg	tgg	gat	gca	cat	tac	aaa	atg	ttt	tgg	ttt	gat	3420
		Ala	Met	Met	Asp	Val	Trp	Asp	Ala	His	Tyr	Lys	Met	Phe	Trp	Phe	Asp	
				625					630					635				
50		cat	cat	ggc	aat	agt	ccc	atg	gcc	agt	ata	cag	gaa	gca	gag	caa	gct	3468
		His	His	Gly	Asn	Ser	Pro	Met	Ala	Ser	Ile	Gln	Glu	Ala	Glu	Gln	Ala	
				640				645					650					
55		gct	gaa	ttt	aaa	aaa	cga	tgg	act	caa	ttt	gtt	aat	gtt	act	ttt	cag	3516
		Ala	Glu	Phe	Lys	Lys	Arg	Trp	Thr	Gln	Phe	Val	Asn	Val	Thr	Phe	Gln	
		655					660					665					670	
60		atg	gaa	ccc	aca	atc	aca	aga	att	gca	tat	gtc	ttt	tat	ggg	cca	tat	3564
		Met	Glu	Pro	Thr	Ile	Thr	Arg	Ile	Ala	Tyr	Val	Phe	Tyr	Gly	Pro	Tyr	
						675					680					685		
		atc	aat	gtc	tcc	agc	tgc	aga	ttt	att	gat	agt	tcc	aat	cct	gga	ctt	3612
		Ile	Asn	Val	Ser	Ser	Cys	Arg	Phe	Ile	Asp	Ser	Ser	Asn	Pro	Gly	Leu	
					690					695					700			

5	cag att tct ctc aat gtc aat aat act gaa cat gtt gtt tct att gta	3660
	Gln Ile Ser Leu Asn Val Asn Asn Thr Glu His Val Val Ser Ile Val	
	705 710 715	
10	act gat tac cat aac ctg aag aca aga ttc aat tat ctg gga ttc ggt	3708
	Thr Asp Tyr His Asn Leu Lys Thr Arg Phe Asn Tyr Leu Gly Phe Gly	
	720 725 730	
15	ggc ttt gcc agt gtg gct gat caa ggc caa ata acc cga ttt ggt ttg	3756
	Gly Phe Ala Ser Val Ala Asp Gln Gly Gln Ile Thr Arg Phe Gly Leu	
	735 740 745 750	
20	ggc act caa gca ata gta aag cct gta aga cat gat agg att att ttc	3804
	Gly Thr Gln Ala Ile Val Lys Pro Val Arg His Asp Arg Ile Ile Phe	
	755 760 765	
25	ccc ttt gga ttt aaa ttt aat ata gca gtt gga tta att ttg tgc att	3852
	Pro Phe Gly Phe Lys Phe Asn Ile Ala Val Gly Leu Ile Leu Cys Ile	
	770 775 780	
30	agc ttg gtg att tta act ttc caa tgg cgt ttt tac ctt tct ttt aga	3900
	Ser Leu Val Ile Leu Thr Phe Gln Trp Arg Phe Tyr Leu Ser Phe Arg	
	785 790 795	
35	aaa cta atg cga tgg ata tta ata ctt gtt att gcc ttg tgg ttt att	3948
	Lys Leu Met Arg Trp Ile Leu Ile Leu Val Ile Ala Leu Trp Phe Ile	
	800 805 810	
40	gag ctt ttg gat gtg tgg agc act tgt agt cag ccc att tgt gca aaa	3996
	Glu Leu Leu Asp Val Trp Ser Thr Cys Ser Gln Pro Ile Cys Ala Lys	
	815 820 825 830	
45	tgg aca agg aca gag gct gag gga agc aag aag tct ttg tct tct gaa	4044
	Trp Thr Arg Thr Glu Ala Glu Gly Ser Lys Lys Ser Leu Ser Ser Glu	
	835 840 845	
50	ggg cac cac atg gat ctt cct gat gtt gtc att acc tca ctt cct ggt	4092
	Gly His His Met Asp Leu Pro Asp Val Val Ile Thr Ser Leu Pro Gly	
	850 855 860	
55	tca gga gct gaa att ctc aaa caa ctt ttt ttc aac agt agt gat ttt	4140
	Ser Gly Ala Glu Ile Leu Lys Gln Leu Phe Phe Asn Ser Ser Asp Phe	
	865 870 875	
60	ctc tac atc agg gtt cct aca gcc tac att gat att cct gaa act gag	4188
	Leu Tyr Ile Arg Val Pro Thr Ala Tyr Ile Asp Ile Pro Glu Thr Glu	
	880 885 890	
65	ttg gaa atc gac tca ttt gta gat gct tgt gaa tgg aag gtg tca gat	4236
	Leu Glu Ile Asp Ser Phe Val Asp Ala Cys Glu Trp Lys Val Ser Asp	
	895 900 905 910	
70	atc cgc agt ggg cat ttt cgt tta ctc cga ggc tgg ttg cag tct tta	4284
	Ile Arg Ser Gly His Phe Arg Leu Leu Arg Gly Trp Leu Gln Ser Leu	
	915 920 925	
75	gtc cag gac aca aaa tta cat ttg caa aac atc cat ctg cat gaa ccc	4332
	Val Gln Asp Thr Lys Leu His Leu Gln Asn Ile His Leu His Glu Pro	
	930 935 940	
80	aat agg ggt aaa ctg gcc caa tat ttt gca atg aat aag gac aaa aaa	4380
	Asn Arg Gly Lys Leu Ala Gln Tyr Phe Ala Met Asn Lys Asp Lys Lys	

	945	950	955	
5	aga aaa ttt aaa agg aga gag tct ttg cca gaa caa aga agt caa atg Arg Lys Phe Lys Arg Arg Glu Ser Leu Pro Glu Gln Arg Ser Gln Met 960 965 970			4428
10	aaa ggc gcc ttt gat aga gat gct gaa tat att agg gct ttg agg aga Lys Gly Ala Phe Asp Arg Asp Ala Glu Tyr Ile Arg Ala Leu Arg Arg 975 980 985 990			4476
15	cac ctg gtt tac tat cca agt gca cgt cct gtg ctc agt tta agc agt His Leu Val Tyr Tyr Pro Ser Ala Arg Pro Val Leu Ser Leu Ser Ser 995 1000 1005			4524
20	gga agc tgg acg tta aag ctt cat ttt ttt cag gaa gtt tta gga gct Gly Ser Trp Thr Leu Lys Leu His Phe Phe Gln Glu Val Leu Gly Ala 1010 1015 1020			4572
25	tcg atg agg gca ttg tac ata gta aga gac cct cgg gca tgg att tat Ser Met Arg Ala Leu Tyr Ile Val Arg Asp Pro Arg Ala Trp Ile Tyr 1025 1030 1035			4620
30	tca atg ttg tac aat agt aaa cca agt ctt tat tct ttg aag aat gta Ser Met Leu Tyr Asn Ser Lys Pro Ser Leu Tyr Ser Leu Lys Asn Val 1040 1045 1050			4668
35	cca gag cat tta gca aaa ttg ttt aaa ata gag gga ggt aaa ggc aaa Pro Glu His Leu Ala Lys Leu Phe Lys Ile Glu Gly Gly Lys Gly Lys 1055 1060 1065 1070			4716
40	tgt aac tta aat tcg ggt tat gct ttc gag tat gaa cca ttg agg aaa Cys Asn Leu Asn Ser Gly Tyr Ala Phe Glu Tyr Glu Pro Leu Arg Lys 1075 1080 1085			4764
45	gaa tta tca aaa tcc aaa tca aat gca gtg tcc ctc ttg tct cac ttg Glu Leu Ser Lys Ser Lys Ser Asn Ala Val Ser Leu Leu Ser His Leu 1090 1095 1100			4812
50	tgg cta gca aat aca gca gca gcc ttg aga ata aat aca gat ttg ctg Trp Leu Ala Asn Thr Ala Ala Ala Leu Arg Ile Asn Thr Asp Leu Leu 1105 1110 1115			4860
55	cct act agc tac cag ctg gtc aag ttt gaa gat att gtg cat ttt cct Pro Thr Ser Tyr Gln Leu Val Lys Phe Glu Asp Ile Val His Phe Pro 1120 1125 1130			4908
60	cag aaa act act gaa agg att ttt gcc ttt ctt gga att cct ttg tct Gln Lys Thr Thr Glu Arg Ile Phe Ala Phe Leu Gly Ile Pro Leu Ser 1135 1140 1145 1150			4956
65	cct gct agt tta aac caa ata ttg ttt gcc acc tct aca aac ctt ttt Pro Ala Ser Leu Asn Gln Ile Leu Phe Ala Thr Ser Thr Asn Leu Phe 1155 1160 1165			5004
70	tac ctt ccc tat gaa ggg gaa ata tca cca act aat act aat gtt tgg Tyr Leu Pro Tyr Glu Gly Glu Ile Ser Pro Thr Asn Thr Asn Val Trp 1170 1175 1180			5052
75	aaa cag aac ttg cct aga gat gaa att aaa cta att gaa aac atc tgc Lys Gln Asn Leu Pro Arg Asp Glu Ile Lys Leu Ile Glu Asn Ile Cys 1185 1190 1195			5100
80	tgg act ctg atg gat cgc cta gga tat cca aag ttt atg gac			5142

	Trp	Thr	Leu	Met	Asp	Arg	Leu	Gly	Tyr	Pro	Lys	Phe	Met	Asp																																																				
	1200						1205					1210																																																						
5	t	a	a	a	t	g	c	t	g	c		a	a	a	t	t	t	g	c	a	c		a	a	a	t	t	t	g	c	a	c	c	a	c	c	c	a		c	t	t	t	g	t	g	g	a	t		5202															
	a	t	g	a	a	t	c	a	g	a		a	g	a	g	t	t	t	g	t		t	a	t	t	c	t	t	t	a	g		t	g	t	g	t	g	t	g		t	g	t	g	t	g	c	a	c	g		c	g	t	g	t	a	t	g	t	g		5262		
	t	t	c	a	g	t	g	t	t	g		t	t	t	g	c	a	c	a	g	a		g	a	g	a	t	t	t	g		t	a	a	a	a	a	t	g		c	a	c	c	a	t	a	t	t	t		g	g	c	c	t	a	g	c	a	g		5322			
10	g	a	t	t	t	a	t	t	t	t		t	a	t	g	t	c	a	t	c	a		c	c	t	c	c	c	t	t	g		c	t	t	t	g	t	t	t	c	t		g	a	a	a	a	t	t	t	g		t	c	t	g	c	t	a	a	a	a		5382	
	a	g	t	t	t	c	t	g	c	t		a	c	a	g	a	g	t	g	t		a	g	a	t	g	a	a	g	t		a	t	a	t	c	a	t	g	g		g	t	c	a	g	g	g	g	g	a		a	t	g	g	g	a	a	a	a	t		5442		
	t	t	t	a	a	g	t	t	t	t		t	g	t	c	t	a	a	c	t		c	c	c	t	t	c	a	t	c	t		g	t	a	a	c	t	g	t	g	c		t	a	a	t	c	t	a	t	c	t		a	g	a	g	a	c	c	t	c	a		5502
15	a	a	c	a	c	t	g	c	t	a		a	a	g	g	c	c	t	t	g		a	a	t	t	g	c	t	g	c	t		t	t	a	c	c	c	a	c	g	c		a	t	c	t	c	t	t	g	c	t		t	t	c	a	a	g	a	t	g	g		5562
	a	c	t	a	c	a	a	a	a	g		t	t	c	c	t	t	a	t	c		t	t	t	t	g	a	a	a	a	g		g	t	c	t	t	c	t	g	a	c		a	c	a	c	t	t	a	t	c	t		t	g	c	a	c	a	a	a	g	a		5622
20	a	a	a	g	a	a	a	a	t		t	t	c	t	t	t	a	c	t		g	t	g	t	t	a	a	t	g		t	t	c	a	g	t	g	a	t	a		t	c	a	c	t	g	a	g	a		a	a	t	g	g	t	g	a	a	a		5682			
	g	c	t	c	t	a	t	c	a		g	a	a	c	t	a	t	a	g		a	t	t	t	c	t	t	c	t	g		g	g	a	a	a	t	a	c	a	g		a	t	g	g	a	a	a	t	a	c		a	g	a	a	t	g	a	a	t	a		5742	
	t	g	t	t	t	t	t	t	t		a	g	g	t	c	g	g	a	a		c	t	g	a	c	t	t	t	a		a	a	g	c	c	t	c	c	t		g	a	a	g	t	t	t	t	t	t		a	c	t	t	a	g	a	a	t		5802				
25	a	t	a	a	g	g	a	a	t	a		a	g	t	c	t	t	t	g	a		c	a	a	t	c	t	g	g	t		g	g	c	a	a	g	g	g	c	t		g	g	t	a	g	a	t	t	a	t		t	t	t	a	g	a	c	a	t	g		5862	
	a	t	t	g	t	c	t	g	t		t	a	a	a	a	c	t	c	t		c	t	t	t	c	a	c	t	t		t	t	a	t	c	c	c	c	c		t	g	g	a	g	c	t	a	c	a		g	c	t	g	t	t	c	g	c		5922				
30	a	t	c	a	c	a	t	c	a		t	c	c	c	a	t	c	c	t		t	c	c	t	t	c	t	g	t		c	a	c	t	g	t	c	a	a	g		c	a	a	a	c	a	a	t	c		a	g	t	a	g	t	t	a	c	t		5982			
	a	a	t	c	g	c	t	g	a		c	t	c	t	c	a	a	t	a		t	g	t	g	g	g	c	a	t		t	t	t	c	c	c	c	c	c	a		g	t	t	g	a	t	t	a	t		t	t	t	g	c	g	t	t	a		6042				
	a	g	a	c	t	g	a	c	a		a	g	a	c	t	t	a	g	a		t	c	a	a	a	t	t	a	t		t	t	t	t	c	t	g	g	a		t	t	a	a	c	a	c	t	c	t		g	t	g	a	c	t	c	a	a	a		6102			
35	g	t	a	g	t	g	c	c	a		t	g	c	a	g	t	g	t	c		t	t	t	t	a	a	a	c	t		g	a	a	c	a	g	a	a	t		t	g	g	a	a	a	a	a	c	t		c	c	t	g	a	c	t	t	a	t		6162			
	c	t	t	g	c	a	t	c	c		t	t	t	g	a	a	t	g	a		t	t	t	a	c	a	g	a	c		g	c	c	a	g	t	g	t	c	t		g	c	a	a	a	a	g	t	t	g		a	a	a	g	c	a	a	a	t	g		6222		
40	g	g	a	g	a	t	g	a	t		t	c	a	g	a	g	c	a	t		c	t	g	t	t	c	c	t	t		t	a	c	c	a	t	c	t	g	c		a	t	c	t	t	a	t	t	a	t		a	a	a	t	g	t	a	g	t	c		6282		
	g	t	c	a	t	a	a	a	g		g	t	g	t	t	t	a	t		t	t	a	t	t	t	t	g	g		a	g	g	c	t	c	t	g	a		a	t	c	a	a	a	a	t	g	c		t	a	c	g	c	c	a	t	t	a		6342				
	t	a	a	g	c	c	a	g	t		g	a	g	t	a	a	t	t	a		a	a	t	g	t	a	t	t	g		a	t	g	a	a	a	c	a	t		a	a	g	g	c	a	g	t	g	t		g	g	a	g	a	c	t	t	g	a		6402			
45	t	g	a	a	a	a	t	c	t		t	g	t	a	c	a	g	a	t		g	c	a	g	t	c	t	t	c		t	c	c	t	g	a	t	g	t		t	c	a	a	a	c	t	g	t	g		g	t	t	c	c	c	c	c	a	a		6462			
	g	c	t	c	t	c	t	a	a		a	c	t	t	g	g	a	a	g		c	t	g	t	c	a	t	t	c		g	a	c	c	t	a	g	a	t		a	a	a	g	t	g	g	t	t	c		t	t	t	c	t	c	a	g	t	a		6522			
50	g	t	t	a	t	t	a	t	t		t	g	t	c	a	a	a	a	t		t	g	c	c	t	c	c	a	g		g	t	g	a	t	a	a	a	g	c		t	c	t	g	t	a	t	a	t	g		t	t	a	g	a	t	t	c	c	a		6582		
	g	c	t	a	a	a	c	c	t		a	c	t	t	g	g	c	t	g		c	a	t	t	t	t	c	t	t		c	c	a	t	t	a	t	a	g	t		g	t	g	a	g	t	g	g	a	g		a	c	t	g	c	c	c	c	c	c		6642		
	c	t	c	c	c	c	a	a	c		t	a	t	t	c	c	t	t	c		c	a	t	a	t	c	t	c	t		a	t	g	a	t	t	g	t	c	c		c	t	c	t	g	t	a	a	t		t	c	a	a	a	a	t	g	a		6702				
55	t	g	a	a	a	t	t	c	a		g	t	g	a	a	t	g	t	a		g	t	t	g	a	g	g	g		c	a	c	t	g	a	a	g	a	c		c	t	g	a	a	t	c	t	a	c		a	c	t	a	g	t	a	a	t	c		6762			
	t	c	a	a	g	a	a	a	g		t	t	a	t	t	c	a	t	t		t	a	t	c	t	c	a	g	a		t	t	a	c	c	g	g	c	a	a		g	c	a	t	a	t	a	a	a	a		t	g	c	t	a	c	t	t	g	g		6822		
60	a	t	a	a	t	a	t	c	t		c	a	t	g	a	a	t	a	t		g	c	a	t	g	c	t	a	c		t	g	g	t	t	g	a	t	a	a		c	a	c	t	a	t	t	t	c	c		a	t	t	a	t	t	g	g	g	c		6882		
	a	g	a	a	t	c	t	c	a		t	g	t	t	t	a	c	t	t		c	a	a	t	c	c	t	a	g		g	a	t	a	t	g	t	g	a	t		c	g	t	g	a	a	t	c	a	g		a	t	c	a	c	a	t	a	t	a		6942		

aaaagtctgg attgtcagta gtattagatc tgatcaaggt aggaattaca attgcatgca 7002
ggtagcaagc aagaaagcag aaactactgt tccctttatt ttaacattgt acagacaata 7062
5 cagaaatgta cctgttggcg gccgggtgca gtggctcacg cctgtaatcc cagcacttcg 7122
ggaggccgag gcgggtggat cacgagggtca ggagatcaag accatcctgg ctaacacggg 7182
10 gaaaccccggt ctctactaaa aaaaaagtac aaaaaattag ccgggcgtgg tggcgggcac 7242
ctgtagtccc agctacacgg gaggtgagg caggagaatg gcatgaacct gggaggcaga 7302
gcttgacagt agtggagatg cgccactgca ctccagcctg ggcgacagag cgagactccg 7362
15 cctcaaaaaa aaaaaaaaaa aagaaaaaaa gaaatgtacc tgttggcagg agaaggccag 7422
atggagtatg tggagtaata gggaaagaag agttacagaa aatgaaaaag aaaatgagtt 7482
acactgagaa tgaatatggg aacacgtcat tgatagcaaa agaaagggtac aggcttacga 7542
20 aaatgatctt tacaatgtat cccagctttc acccccacat ggcaatgcag agttgtattt 7602
acttgtttct gtactcacct actcccaccc caagggaaga ttttagacat gaaccctact 7662
25 atttagttat tctaaaatag aaagtttgct ggagaaagcg tctactcaca gattgttctg 7722
taaggaatgt tatgtatggg tgagcgggtg acacatccat tgggtatgta tgcattgatg 7782
gggtgcctgag acccctgcct tagaaacaga attcctaagg ggattgactc tcccagcatg 7842
30 tttccaggtc ctgcaccctt aggtgatctt aggaaaattt taaatagctt ctactcttat 7902
ttttgttctt tgaataaatt aaaagaggga ttatcactat ctgatacttc tgaagaaac 7962
35 acttacaaaa tttcttatct gtaaaatccg tctttttcta cattaacttc ccaaacata 8022
ggcctaattg agataattgc ttttattata ataataggat tgaaatttta aaattttgaa 8082
aggacttatt aattttgctg acaaaagtga agtaacaaat ataatgataa ttggcttttt 8142
40 aaattttcaa acaacataga tttactcaag atgaaataaa aaggccatat tcagagttga 8202
atttaatgaa aactcagagg aaataggaaa atctgctcag gagaaagaag ctaaactctg 8262
45 atagatttag tttgtagaat ttaatttaaa atttaaattt taacaaagtg atgacacaac 8322
aatatgtacg tttagggtgtg gacacaaaaa tattagacat ttgattgtcc ttttacatag 8382
agaataacta ataaatgcct gacaagaatg ggacaatcct tccttgatc aaaattccca 8442
50 ggtcttgcta cattgccctc tgcaaatgta ttcaaagaag aacctcctcc accacttact 8502
tttggttggc ataattgttc agcaacgatt tctgtacatc accaagtatc tttggcattc 8562
55 ttggtataca agtatatca caattttaag tgagtaata ttaatgataa tttttgaatt 8622
gctttgtttg gcttgattaa ctttgatcag aaatagaaac gttttcattt gttgatttag 8682
gaaaaagcat aaatagaatg cagtataaca ccacttccaa aggtaaggat acctaacatt 8742
60 cttttttttt tttttttttt ttttggggat ggagtctcac tttgttggcc aggctggagt 8802
gcagtgggtc gatctcggct cactgcaacc tccgcctacc gggttcaagt gattctccta 8862

5 cgtcagcctc ctgaatagct gggattacag gtgcacgcca ccatgcttgg ctcatttttg 8922
 tatttttagt agtgacagcg tttcaccaca ttggtcaggc tggntctcaa tctcttgacc 8982
 tgggtgatctg cccacctggg cctcccaaaa tgctgggatt acaggcatga gccaccacac 9042
 ctggcaaggg tacctgacat tctaagatat caagacactt aatatgtggg ctattagctg 9102
 10 cttattttaa tgttgaccaa attgtctgat atatctgatt aatcatgatt tcacttcatt 9162
 tcggaagaaa aattatccat atcattttta aagacgcaaa tgactttgga tttttgcata 9222
 gagtacaata gacattcaa acaatagatt ctaacattct ctgaaacact tgagatgttt 9282
 15 gagctacat ttatatgggt tatttatatt tagtctaagt aacacataca tgtttaattg 9342
 attctgtttt catggataga ttcaactaag tcttccaagc aattaatttt ttgttcgtcg 9402
 20 tcgttttttc ttcatacggt atctagttat gcagcactgg aaacagactg aagatcataa 9462
 accagtttta tcagacctat gtgtaataag actcctgtta atacaaaaat aaaaagctaa 9522
 aagcaa 9528

25
 <210> 2
 <211> 1212
 <212> PRT
 30 <213> Homo sapiens
 <220>
 <221> Amino acid sequence encoding Human NCAG1 protein

35 <400> 2
 Met Ala Leu Met Phe Thr Gly His Leu Leu Phe Leu Ala Leu Leu Met
 1 5 10 15
 Phe Ala Phe Ser Thr Phe Glu Glu Ser Val Ser Asn Tyr Ser Glu Trp
 40 20 25 30
 Ala Val Phe Thr Asp Asp Ile Asp Gln Phe Lys Thr Gln Lys Val Gln
 35 40 45
 45 Asp Phe Arg Pro Asn Gln Lys Leu Lys Lys Ser Met Leu His Pro Ser
 50 55 60
 Leu Tyr Phe Asp Ala Gly Glu Ile Gln Ala Met Arg Gln Lys Ser Arg
 50 65 70 75 80
 Ala Ser His Leu His Leu Phe Arg Ala Ile Arg Ser Ala Val Thr Val
 85 90 95
 55 Met Leu Ser Asn Pro Thr Tyr Tyr Leu Pro Pro Pro Lys His Ala Asp
 100 105 110
 Phe Ala Ala Lys Trp Asn Glu Ile Tyr Gly Asn Asn Leu Pro Pro Leu
 115 120 125
 60 Ala Leu Tyr Cys Leu Leu Cys Pro Glu Asp Lys Val Ala Phe Glu Phe
 130 135 140
 Val Leu Glu Tyr Met Asp Arg Met Val Gly Tyr Lys Asp Trp Leu Val

10/22

	145		150		155		160
	Glu Asn Ala Pro Gly Asp Glu Val Pro Ile Gly His Ser Leu Thr Gly						
		165			170		175
5	Phe Ala Thr Ala Phe Asp Phe Leu Tyr Asn Leu Leu Asp Asn His Arg						
		180		185		190	
10	Arg Gln Lys Tyr Leu Glu Lys Ile Trp Val Ile Thr Glu Glu Met Tyr						
		195		200		205	
	Glu Tyr Ser Lys Val Arg Ser Trp Gly Lys Gln Leu Leu His Asn His						
		210		215		220	
15	Gln Ala Thr Asn Met Ile Ala Leu Leu Thr Gly Ala Leu Val Thr Gly						
			230		235		240
	Val Asp Lys Gly Ser Lys Ala Asn Ile Trp Lys Gln Ala Val Val Asp						
		245		250		255	
20	Val Met Glu Lys Thr Met Phe Leu Leu Asn His Ile Val Asp Gly Ser						
		260		265		270	
	Leu Asp Glu Gly Val Ala Tyr Gly Ser Tyr Thr Ala Lys Ser Val Thr						
25		275		280		285	
	Gln Tyr Val Phe Leu Ala Gln Arg His Phe Asn Ile Asn Asn Leu Asp						
		290		295		300	
30	Asn Asn Trp Leu Lys Met His Phe Trp Phe Tyr Tyr Ala Thr Leu Leu						
			310		315		320
	Pro Gly Phe Gln Arg Thr Val Gly Ile Ala Asp Ser Asn Tyr Asn Trp						
		325		330		335	
35	Phe Tyr Gly Pro Glu Ser Gln Leu Val Phe Leu Asp Lys Phe Ile Leu						
		340		345		350	
	Lys Asn Gly Ala Gly Asn Trp Leu Ala Gln Gln Ile Arg Lys His Arg						
40		355		360		365	
	Pro Lys Asp Gly Pro Met Val Pro Ser Thr Ala Gln Arg Trp Ser Thr						
		370		375		380	
45	Leu His Thr Glu Tyr Ile Trp Tyr Asp Pro Gln Leu Thr Pro Gln Pro						
			390		395		400
	Pro Ala Asp Tyr Gly Thr Ala Lys Ile His Thr Phe Pro Asn Trp Gly						
		405		410		415	
50	Val Val Thr Tyr Gly Ala Gly Leu Pro Asn Thr Gln Thr Asn Thr Phe						
		420		425		430	
	Val Ser Phe Lys Ser Gly Lys Leu Gly Gly Arg Ala Val Tyr Asp Ile						
55		435		440		445	
	Val His Phe Gln Pro Tyr Ser Trp Ile Asp Gly Trp Arg Ser Phe Asn						
		450		455		460	
60	Pro Gly His Glu His Pro Asp Gln Asn Ser Phe Thr Phe Ala Pro Asn						
			470		475		480
	Gly Gln Val Phe Val Ser Glu Ala Leu Tyr Gly Pro Lys Leu Ser His						

11/22

	485					490					495					
	Leu	Asn	Asn	Val	Leu	Val	Phe	Ala	Pro	Ser	Pro	Ser	Ser	Gln	Cys	Asn
				500					505					510		
5	Lys	Pro	Trp	Glu	Gly	Gln	Leu	Gly	Glu	Cys	Ala	Gln	Trp	Leu	Lys	Trp
			515					520					525			
10	Thr	Gly	Glu	Glu	Val	Gly	Asp	Ala	Ala	Gly	Glu	Ile	Ile	Thr	Ala	Ser
		530					535					540				
	Gln	His	Gly	Glu	Met	Val	Phe	Val	Ser	Gly	Glu	Ala	Val	Ser	Ala	Tyr
	545					550					555					560
15	Ser	Ser	Ala	Met	Arg	Leu	Lys	Ser	Val	Tyr	Arg	Ala	Leu	Leu	Leu	Leu
					565					570						575
	Asn	Ser	Gln	Thr	Leu	Leu	Val	Val	Asp	His	Ile	Glu	Arg	Gln	Glu	Asp
20				580					585					590		
	Ser	Pro	Ile	Asn	Ser	Val	Ser	Ala	Phe	Phe	His	Asn	Leu	Asp	Ile	Asp
			595					600					605			
25	Phe	Lys	Tyr	Ile	Pro	Tyr	Lys	Phe	Met	Asn	Arg	Tyr	Asn	Gly	Ala	Met
		610					615					620				
	Met	Asp	Val	Trp	Asp	Ala	His	Tyr	Lys	Met	Phe	Trp	Phe	Asp	His	His
	625					630					635					640
30	Gly	Asn	Ser	Pro	Met	Ala	Ser	Ile	Gln	Glu	Ala	Glu	Gln	Ala	Ala	Glu
					645					650						655
	Phe	Lys	Lys	Arg	Trp	Thr	Gln	Phe	Val	Asn	Val	Thr	Phe	Gln	Met	Glu
35				660					665					670		
	Pro	Thr	Ile	Thr	Arg	Ile	Ala	Tyr	Val	Phe	Tyr	Gly	Pro	Tyr	Ile	Asn
			675					680					685			
40	Val	Ser	Ser	Cys	Arg	Phe	Ile	Asp	Ser	Ser	Asn	Pro	Gly	Leu	Gln	Ile
		690					695					700				
	Ser	Leu	Asn	Val	Asn	Asn	Thr	Glu	His	Val	Val	Ser	Ile	Val	Thr	Asp
	705					710					715					720
45	Tyr	His	Asn	Leu	Lys	Thr	Arg	Phe	Asn	Tyr	Leu	Gly	Phe	Gly	Gly	Phe
					725					730					735	
	Ala	Ser	Val	Ala	Asp	Gln	Gly	Gln	Ile	Thr	Arg	Phe	Gly	Leu	Gly	Thr
				740					745					750		
50	Gln	Ala	Ile	Val	Lys	Pro	Val	Arg	His	Asp	Arg	Ile	Ile	Phe	Pro	Phe
			755					760					765			
55	Gly	Phe	Lys	Phe	Asn	Ile	Ala	Val	Gly	Leu	Ile	Leu	Cys	Ile	Ser	Leu
		770					775					780				
	Val	Ile	Leu	Thr	Phe	Gln	Trp	Arg	Phe	Tyr	Leu	Ser	Phe	Arg	Lys	Leu
	785					790					795					800
60	Met	Arg	Trp	Ile	Leu	Ile	Leu	Val	Ile	Ala	Leu	Trp	Phe	Ile	Glu	Leu
				805						810					815	
	Leu	Asp	Val	Trp	Ser	Thr	Cys	Ser	Gln	Pro	Ile	Cys	Ala	Lys	Trp	Thr

12/22

	820	825	830
	Arg Thr Glu Ala Glu Gly Ser Lys Lys Ser Leu Ser Ser Glu Gly His		
	835	840	845
5	His Met Asp Leu Pro Asp Val Val Ile Thr Ser Leu Pro Gly Ser Gly		
	850	855	860
10	Ala Glu Ile Leu Lys Gln Leu Phe Phe Asn Ser Ser Asp Phe Leu Tyr		
	865	870	875
	Ile Arg Val Pro Thr Ala Tyr Ile Asp Ile Pro Glu Thr Glu Leu Glu		
	885	890	895
15	Ile Asp Ser Phe Val Asp Ala Cys Glu Trp Lys Val Ser Asp Ile Arg		
	900	905	910
	Ser Gly His Phe Arg Leu Leu Arg Gly Trp Leu Gln Ser Leu Val Gln		
	915	920	925
20	Asp Thr Lys Leu His Leu Gln Asn Ile His Leu His Glu Pro Asn Arg		
	930	935	940
25	Gly Lys Leu Ala Gln Tyr Phe Ala Met Asn Lys Asp Lys Lys Arg Lys		
	945	950	955
	Phe Lys Arg Arg Glu Ser Leu Pro Glu Gln Arg Ser Gln Met Lys Gly		
	965	970	975
30	Ala Phe Asp Arg Asp Ala Glu Tyr Ile Arg Ala Leu Arg Arg His Leu		
	980	985	990
	Val Tyr Tyr Pro Ser Ala Arg Pro Val Leu Ser Leu Ser Ser Gly Ser		
	995	1000	1005
35	Trp Thr Leu Lys Leu His Phe Phe Gln Glu Val Leu Gly Ala Ser Met		
	1010	1015	1020
40	Arg Ala Leu Tyr Ile Val Arg Asp Pro Arg Ala Trp Ile Tyr Ser Met		
	1025	1030	1035
	Leu Tyr Asn Ser Lys Pro Ser Leu Tyr Ser Leu Lys Asn Val Pro Glu		
	1045	1050	1055
45	His Leu Ala Lys Leu Phe Lys Ile Glu Gly Gly Lys Gly Lys Cys Asn		
	1060	1065	1070
	Leu Asn Ser Gly Tyr Ala Phe Glu Tyr Glu Pro Leu Arg Lys Glu Leu		
	1075	1080	1085
50	Ser Lys Ser Lys Ser Asn Ala Val Ser Leu Leu Ser His Leu Trp Leu		
	1090	1095	1100
55	Ala Asn Thr Ala Ala Ala Leu Arg Ile Asn Thr Asp Leu Leu Pro Thr		
	1105	1110	1115
	Ser Tyr Gln Leu Val Lys Phe Glu Asp Ile Val His Phe Pro Gln Lys		
	1125	1130	1135
60	Thr Thr Glu Arg Ile Phe Ala Phe Leu Gly Ile Pro Leu Ser Pro Ala		
	1140	1145	1150
	Ser Leu Asn Gln Ile Leu Phe Ala Thr Ser Thr Asn Leu Phe Tyr Leu		

1155 1160 1165
 Pro Tyr Glu Gly Glu Ile Ser Pro Thr Asn Thr Asn Val Trp Lys Gln
 1170 1175 1180
 5 Asn Leu Pro Arg Asp Glu Ile Lys Leu Ile Glu Asn Ile Cys Trp Thr
 185 1190 1195 1200
 10 Leu Met Asp Arg Leu Gly Tyr Pro Lys Phe Met Asp
 1205 1210

 15 <210> 3
 <211> 5092
 <212> DNA
 <213> Mus sp.

 20 <220>
 <221> CDS encoding the mouse NCAG1 protein
 <222> (501)..(4121)

 <400> 3
 25 tctgagaatg acagtacttt atcatcttct tttggggaac atacagaaac ataccattta 60
 tgtgtggttaa gttaatcact acagatgggt tcttgtgcta cgtgggtcaaa tggcttcatt 120
 tgaatttttg aatttttaaaa aattttttct ttttcacatg ttaattagat ttacacacag 180
 30 ggagtaaagt ttggatttgt tgtattttct gactagacca ctgttttctg tgcattggag 240
 acattggagg cattaatatt ccttgaaatt ttattttatt ggaagcaaac ctgtgccagg 300
 gacacagaca tgctatataa tttcctaact tttcttgctt tgaataagct gaatgtcacc 360
 35 tggatttcac agcctatgag gtatagtctg ttttttggtt ttgttttttt gctacatctt 420
 taatatataa tttacaataa ccagatggga aacctgtgc ttaacacata tgcctaagga 480
 40 aaagatcttc cccatggatc atg gcg ttt atg ttt aca gaa cat tta cta ttt 533
 Met Ala Phe Met Phe Thr Glu His Leu Leu Phe
 1 5 10

 45 tta aca ttg atg atg tgt agt ttt tct act tgt gaa gaa tct gtg agc 581
 Leu Thr Leu Met Met Cys Ser Phe Ser Thr Cys Glu Glu Ser Val Ser
 15 20 25

 aat tat tct gaa tgg gca gtt ttc aca gac gat ata caa tgg ctt aag 629
 Asn Tyr Ser Glu Trp Ala Val Phe Thr Asp Asp Ile Gln Trp Leu Lys
 30 35 40
 50 tca cag aaa ata caa gat ttc aaa ctc aac cga aga ctt cat cca aat 677
 Ser Gln Lys Ile Gln Asp Phe Lys Leu Asn Arg Arg Leu His Pro Asn
 45 50 55
 55 tta tat ttt gat gct gga gat ata caa aca ttg aaa caa aag tct cgt 725
 Leu Tyr Phe Asp Ala Gly Asp Ile Gln Thr Leu Lys Gln Lys Ser Arg
 60 65 70 75
 60 aca agc cat ttg cat att ttt aga gct atc aaa agt gca gtg aca att 773
 Thr Ser His Leu His Ile Phe Arg Ala Ile Lys Ser Ala Val Thr Ile
 80 85 90

	atg ctg tcc aat cca tca tac tac cta cct cca ccc aag cat gct gag	821
	Met Leu Ser Asn Pro Ser Tyr Tyr Leu Pro Pro Pro Lys His Ala Glu	
	95 100 105	
5	ttt gct gcc aag tgg aat gaa att tat ggt aat aat ctt cct cct tta	869
	Phe Ala Ala Lys Trp Asn Glu Ile Tyr Gly Asn Asn Leu Pro Pro Leu	
	110 115 120	
10	gca ttg tat tgt tta tta tgc cca gaa gac aag gtt gcc ttt gaa ttt	917
	Ala Leu Tyr Cys Leu Leu Cys Pro Glu Asp Lys Val Ala Phe Glu Phe	
	125 130 135	
15	gtt atg gaa tac atg gat cgg atg gtt agc tac aaa gac tgg cta gtt	965
	Val Met Glu Tyr Met Asp Arg Met Val Ser Tyr Lys Asp Trp Leu Val	
	140 145 150 155	
20	gag aat gca cca ggg gat gag gtt cca gtt ggc cat tct tta aca ggt	1013
	Glu Asn Ala Pro Gly Asp Glu Val Pro Val Gly His Ser Leu Thr Gly	
	160 165 170	
25	ttt gcc act gcc ttt gac ttt tta tat aat cta tta ggt aat cag cgt	1061
	Phe Ala Thr Ala Phe Asp Phe Leu Tyr Asn Leu Leu Gly Asn Gln Arg	
	175 180 185	
30	aaa caa aaa tac cta gaa aaa att tgg att gtt act gag gaa atg tat	1109
	Lys Gln Lys Tyr Leu Glu Lys Ile Trp Ile Val Thr Glu Glu Met Tyr	
	190 195 200	
35	gaa tat tcc aag att cga tca tgg ggc aaa caa ctt ctt cat aac cat	1157
	Glu Tyr Ser Lys Ile Arg Ser Trp Gly Lys Gln Leu Leu His Asn His	
	205 210 215	
40	caa gct aca aat atg ata gct tta ctc ata ggg gcc ttg gtt act gga	1205
	Gln Ala Thr Asn Met Ile Ala Leu Leu Ile Gly Ala Leu Val Thr Gly	
	220 225 230 235	
45	gta gat aaa gga tct aaa gca aac ata tgg aaa caa gtt gtt gtt gat	1253
	Val Asp Lys Gly Ser Lys Ala Asn Ile Trp Lys Gln Val Val Val Asp	
	240 245 250	
50	gtg atg gaa aag act atg ttt ctc ttg aag cat att gta gat ggc tca	1301
	Val Met Glu Lys Thr Met Phe Leu Leu Lys His Ile Val Asp Gly Ser	
	255 260 265	
55	ttg gat gaa ggt gtg gcc tat gga agc tat acc tca aaa tca gtt aca	1349
	Leu Asp Glu Gly Val Ala Tyr Gly Ser Tyr Thr Ser Lys Ser Val Thr	
	270 275 280	
60	cag tat gtt ttt ttg gca caa cgc cat ttt aac atc aac aac ttt gat	1397
	Gln Tyr Val Phe Leu Ala Gln Arg His Phe Asn Ile Asn Asn Phe Asp	
	285 290 295	
65	aat aac tgg cta aaa atg cat ttt tgg ttt tat tat gct aca ctt ttg	1445
	Asn Asn Trp Leu Lys Met His Phe Trp Phe Tyr Tyr Ala Thr Leu Leu	
	300 305 310 315	
70	cca ggc tat caa aga act gta ggc ata gca gat tcc aat tat aat tgg	1493
	Pro Gly Tyr Gln Arg Thr Val Gly Ile Ala Asp Ser Asn Tyr Asn Trp	
	320 325 330	
75	ttt tat ggt cca gag agc cag cta gtt ttc ttg gat aag ttc att tta	1541
	Phe Tyr Gly Pro Glu Ser Gln Leu Val Phe Leu Asp Lys Phe Ile Leu	
	335 340 345	

5	cag aat gga gct gga aat tgg tta gct cag caa att aga aag cat cga	1589
	Gln Asn Gly Ala Gly Asn Trp Leu Ala Gln Gln Ile Arg Lys His Arg	
	350 355 360	
10	cct aag gat gga cca atg gtt cct tcc act gct cag cgg tgg agt act	1637
	Pro Lys Asp Gly Pro Met Val Pro Ser Thr Ala Gln Arg Trp Ser Thr	
	365 370 375	
15	ctt cat act gaa tac atc tgg tat gat cca aca ctc acc cca cag cct	1685
	Leu His Thr Glu Tyr Ile Trp Tyr Asp Pro Thr Leu Thr Pro Gln Pro	
	380 385 390 395	
20	cct gtt gat ttt ggc act gca aaa atg cac aca ttt cct aac tgg ggt	1733
	Pro Val Asp Phe Gly Thr Ala Lys Met His Thr Phe Pro Asn Trp Gly	
	400 405 410	
25	gtc gtg act tat ggg ggt ggg ctg cca aac acc cag acc aat acc ttt	1781
	Val Val Thr Tyr Gly Gly Gly Leu Pro Asn Thr Gln Thr Asn Thr Phe	
	415 420 425	
30	gtg tct ttt aaa tct ggg aaa ctg gga gga cga gct gtg tat gac ata	1829
	Val Ser Phe Lys Ser Gly Lys Leu Gly Gly Arg Ala Val Tyr Asp Ile	
	430 435 440	
35	gtt cac ttt cag cca tat tcc tgg att gat gga tgg aga agc ttt aac	1877
	Val His Phe Gln Pro Tyr Ser Trp Ile Asp Gly Trp Arg Ser Phe Asn	
	445 450 455	
40	cca gga cat gaa cat cca gat caa aat tca ttt act ttc gct cct aat	1925
	Pro Gly His Glu His Pro Asp Gln Asn Ser Phe Thr Phe Ala Pro Asn	
	460 465 470 475	
45	ggg cag gta ttc gtt tct gag gct ctt tat gga cca aaa ttg agc cac	1973
	Gly Gln Val Phe Val Ser Glu Ala Leu Tyr Gly Pro Lys Leu Ser His	
	480 485 490	
50	ctt aac aac gta ttg gtg ttt gcc cca tca cca tca agt caa tgt aat	2021
	Leu Asn Asn Val Leu Val Phe Ala Pro Ser Pro Ser Ser Gln Cys Asn	
	495 500 505	
55	cag ccc tgg gaa ggt caa ctg gga gaa tgt gca cag tgg ctc aag tgg	2069
	Gln Pro Trp Glu Gly Gln Leu Gly Glu Cys Ala Gln Trp Leu Lys Trp	
	510 515 520	
60	act ggg gaa gag gtt ggt gat gca gct ggg gaa gtt att act gct gct	2117
	Thr Gly Glu Glu Val Gly Asp Ala Ala Gly Glu Val Ile Thr Ala Ala	
	525 530 535	
65	caa cat ggt gat agg atg ttt gtg agt ggg gaa gca gtg tct gct tat	2165
	Gln His Gly Asp Arg Met Phe Val Ser Gly Glu Ala Val Ser Ala Tyr	
	540 545 550 555	
70	tct tct gcc atg aga ctg aaa agt gtc tat cgt gct tta ctt ctt tta	2213
	Ser Ser Ala Met Arg Leu Lys Ser Val Tyr Arg Ala Leu Leu Leu Leu	
	560 565 570	
75	aat tca caa act ctg ctt gtt gtc gat cat att gaa agg caa gaa act	2261
	Asn Ser Gln Thr Leu Leu Val Val Asp His Ile Glu Arg Gln Glu Thr	
	575 580 585	
80	tcc cca ata aat tct gtc agt gcc ttc ttt cat aat ttg gat att gat	2309
	Ser Pro Ile Asn Ser Val Ser Ala Phe Phe His Asn Leu Asp Ile Asp	

	590	595	600	
5	ttt aaa tac atc cca tac aag ttt atg aat aga tat aat ggt gcc atg Phe Lys Tyr Ile Pro Tyr Lys Phe Met Asn Arg Tyr Asn Gly Ala Met 605 610 615	2357		
10	atg gat gtg tgg gat gca cac tat aaa atg ttt tgg ttt gat cac cat Met Asp Val Trp Asp Ala His Tyr Lys Met Phe Trp Phe Asp His His 620 625 630 635	2405		
	ggc aac agt cct gtg gct aat ata cag gaa gca gaa cag gct gct gaa Gly Asn Ser Pro Val Ala Asn Ile Gln Glu Ala Glu Gln Ala Ala Glu 640 645 650	2453		
15	ttt aag aaa cgg tgg aca cag ttt gtt aat gtt aca ttt cat atg gaa Phe Lys Lys Arg Trp Thr Gln Phe Val Asn Val Thr Phe His Met Glu 655 660 665	2501		
20	tcc aca atc aca aga att gct tat gta ttt tat ggg cca tat gtc aat Ser Thr Ile Thr Arg Ile Ala Tyr Val Phe Tyr Gly Pro Tyr Val Asn 670 675 680	2549		
25	gtt tcc agc tgc aga ttt att gat agt tcc agt tct gga ctt cag att Val Ser Ser Cys Arg Phe Ile Asp Ser Ser Ser Ser Gly Leu Gln Ile 685 690 695	2597		
30	tct tta cat gtc aac agt act gaa cat agt gtg tct gtt gta act gac Ser Leu His Val Asn Ser Thr Glu His Ser Val Ser Val Val Thr Asp 700 705 710 715	2645		
	tat caa aac ctt aaa agc aga ttc agt tac ctg gga ttt ggt ggt ttt Tyr Gln Asn Leu Lys Ser Arg Phe Ser Tyr Leu Gly Phe Gly Gly Phe 720 725 730	2693		
35	gcc agt gtg gct aat caa gga cag ata acc aga ttt ggt ttg ggt act Ala Ser Val Ala Asn Gln Gly Gln Ile Thr Arg Phe Gly Leu Gly Thr 735 740 745	2741		
40	caa gaa ata gta aac cct gta aga cat gat aaa gtt aat ttc ccc ttt Gln Glu Ile Val Asn Pro Val Arg His Asp Lys Val Asn Phe Pro Phe 750 755 760	2789		
45	ggg ttt aaa ttt aat ata gca gtt gga ttc att ttg tgt att agt ttg Gly Phe Lys Phe Asn Ile Ala Val Gly Phe Ile Leu Cys Ile Ser Leu 765 770 775	2837		
50	gtt att tta act ttt caa tgg cgg ttt tac ctt tcc ttt aga aag cta Val Ile Leu Thr Phe Gln Trp Arg Phe Tyr Leu Ser Phe Arg Lys Leu 780 785 790 795	2885		
	atg cgc tgt gta tta ata ctt gtt att gcc ttg tgg ttt att gag ctt Met Arg Cys Val Leu Ile Leu Val Ile Ala Leu Trp Phe Ile Glu Leu 800 805 810	2933		
55	ctg gat gta tgg agt aca tgc act cag ccc atc tgt gca aaa tgg aca Leu Asp Val Trp Ser Thr Cys Thr Gln Pro Ile Cys Ala Lys Trp Thr 815 820 825	2981		
60	agg act gaa gct aag gca aat gag aag gtc atg att tct gaa ggg cat Arg Thr Glu Ala Lys Ala Asn Glu Lys Val Met Ile Ser Glu Gly His 830 835 840	3029		
	cat gtg gat ctt cct aat gtt att att acc tca ctc cct ggt tca gga	3077		

	His	Val	Asp	Leu	Pro	Asn	Val	Ile	Ile	Thr	Ser	Leu	Pro	Gly	Ser	Gly	
	845						850					855					
5	gct	gaa	att	ctc	aaa	cag	ctt	ttt	ttc	aac	agc	agt	gat	ttt	ctc	tac	3125
	Ala	Glu	Ile	Leu	Lys	Gln	Leu	Phe	Phe	Asn	Ser	Ser	Asp	Phe	Leu	Tyr	
	860					865					870					875	
10	atc	aga	att	cct	aca	gcc	tac	atg	gat	atc	cct	gaa	act	gaa	ttt	gaa	3173
	Ile	Arg	Ile	Pro	Thr	Ala	Tyr	Met	Asp	Ile	Pro	Glu	Thr	Glu	Phe	Glu	
					880					885						890	
15	att	gac	tca	ttt	gta	gat	gct	tgt	gag	tgg	aaa	gta	tca	gat	atc	cgc	3221
	Ile	Asp	Ser	Phe	Val	Asp	Ala	Cys	Glu	Trp	Lys	Val	Ser	Asp	Ile	Arg	
				895					900							905	
20	agt	ggg	cac	ttt	cat	ctt	ctt	cga	ggg	tgg	ctg	cag	tct	ttg	gtc	cag	3269
	Ser	Gly	His	Phe	His	Leu	Leu	Arg	Gly	Trp	Leu	Gln	Ser	Leu	Val	Gln	
			910					915						920			
25	gat	aca	aaa	ctt	cac	ttg	caa	aac	atc	cat	cta	cat	gaa	acc	agt	agg	3317
	Asp	Thr	Lys	Leu	His	Leu	Gln	Asn	Ile	His	Leu	His	Glu	Thr	Ser	Arg	
		925					930						935				
30	agt	aaa	ctg	gcc	caa	tat	ttt	aca	act	aat	aag	gac	aaa	aag	cga	aaa	3365
	Ser	Lys	Leu	Ala	Gln	Tyr	Phe	Thr	Thr	Asn	Lys	Asp	Lys	Lys	Arg	Lys	
		940				945					950					955	
35	tta	aaa	aga	agg	gag	tct	ttg	caa	gat	caa	aga	agt	aga	ata	aaa	gga	3413
	Leu	Lys	Arg	Arg	Glu	Ser	Leu	Gln	Asp	Gln	Arg	Ser	Arg	Ile	Lys	Gly	
					960					965						970	
40	cca	ttt	gat	aga	gat	gct	gaa	tat	att	agg	gct	tta	aga	aga	cac	ctt	3461
	Pro	Phe	Asp	Arg	Asp	Ala	Glu	Tyr	Ile	Arg	Ala	Leu	Arg	Arg	His	Leu	
				975					980						985		
45	gtt	tat	tac	cca	agt	gca	cgt	cct	gtg	ctc	agc	tta	agt	agt	ggg	agc	3509
	Val	Tyr	Tyr	Pro	Ser	Ala	Arg	Pro	Val	Leu	Ser	Leu	Ser	Ser	Gly	Ser	
			990					995					1000				
50	tgg	aca	ttg	aag	ctt	cat	ttt	ttt	cag	gaa	gtt	tta	gga	act	tca	atg	3557
	Trp	Thr	Leu	Lys	Leu	His	Phe	Phe	Gln	Glu	Val	Leu	Gly	Thr	Ser	Met	
		1005					1010						1015				
55	cgg	gca	ttg	tac	ata	gta	aga	gac	cct	cga	gct	tgg	atc	tat	tca	gtg	3605
	Arg	Ala	Leu	Tyr	Ile	Val	Arg	Asp	Pro	Arg	Ala	Trp	Ile	Tyr	Ser	Val	
		1020				1025					1030					1035	
60	cta	tat	ggg	agt	aaa	cca	agt	ctt	tat	tct	ttg	aag	aat	gta	cca	gag	3653
	Leu	Tyr	Gly	Ser	Lys	Pro	Ser	Leu	Tyr	Ser	Leu	Lys	Asn	Val	Pro	Glu	
					1040					1045					1050		
65	cac	tta	gca	aaa	ttg	ttt	aaa	ata	gag	gaa	ggg	aaa	agc	aaa	tgt	aat	3701
	His	Leu	Ala	Lys	Leu	Phe	Lys	Ile	Glu	Gly	Lys	Ser	Lys	Cys	Asn		
				1055					1060					1065			
70	tcg	aat	tct	ggc	tat	gct	ttt	gag	tat	gaa	tca	ctg	aag	aaa	gaa	tta	3749
	Ser	Asn	Ser	Gly	Tyr	Ala	Phe	Glu	Tyr	Glu	Ser	Leu	Lys	Lys	Glu	Leu	
			1070					1075					1080				
75	gaa	ata	tcc	caa	tca	aat	gct	atc	tcc	tta	tta	tct	cat	ttg	tgg	gta	3797
	Glu	Ile	Ser	Gln	Ser	Asn	Ala	Ile	Ser	Leu	Leu	Ser	His	Leu	Trp	Val	
		1085					1090						1095				

gca aac act gca gca gcc ttg aga ata aat aca gat ttg ctg cct acc 3845
 Ala Asn Thr Ala Ala Ala Leu Arg Ile Asn Thr Asp Leu Leu Pro Thr
 1100 1105 1110 1115
 5 aat tac cat ctg gtc aag ttt gaa gat att gtt cat ttt cct cag aag 3893
 Asn Tyr His Leu Val Lys Phe Glu Asp Ile Val His Phe Pro Gln Lys
 1120 1125 1130
 10 act act gaa agg att ttt gct ttc ctt ggc att cct ttg tct cct gct 3941
 Thr Thr Glu Arg Ile Phe Ala Phe Leu Gly Ile Pro Leu Ser Pro Ala
 1135 1140 1145
 15 agt tta aac caa atg cta ttt gcc act tcc aca aac ctt ttt tat ctt 3989
 Ser Leu Asn Gln Met Leu Phe Ala Thr Ser Thr Asn Leu Phe Tyr Leu
 1150 1155 1160
 20 cca tat gag ggg gaa ata tca cca tct aat act aat att tgg aaa aca 4037
 Pro Tyr Glu Gly Glu Ile Ser Pro Ser Asn Thr Asn Ile Trp Lys Thr
 1165 1170 1175
 aac ttg cct aga gat gaa att aaa cta att gaa aac att tgc tgg aca 4085
 Asn Leu Pro Arg Asp Glu Ile Lys Leu Ile Glu Asn Ile Cys Trp Thr
 1180 1185 1190 1195
 25 ctg atg gat cat cta gga tat cca aag ttt atg gac taaatgctgc 4131
 Leu Met Asp His Leu Gly Tyr Pro Lys Phe Met Asp
 1200 1205
 30 aggtcggcaa aatttgact aatgtgtccc aacctacttt gtggatatga actagaaaac 4191
 tttgtttatt cttgtacatg tatgtatgtg tgtagagtga gtgcgtgtgt ccagtatgtt 4251
 atttgcacag agatattttc aaaataggca ccatatttgg cctagcagga tttattttta 4311
 35 tgttaccact tttcttgcct ttgtttctga atttttttct gctaaaatgt ttctgctaca 4371
 gaggtatata ttctggggtt ctgaaatatg ggggttttaac ggactttaac tcaacttctt 4431
 40 tggaaactat ttatctatct taggacctca aacactacaa acggccttgc aattgctgct 4491
 gtatctagtc atctctcgct cttaatatgg actacaaaac tttatgtttt gaaaacgtct 4551
 aacatttacc ttgcacacaa aaacgagaaa taaaaaaaca aaaattattt tacgttgtat 4611
 45 agtgtttatt gaaatcactt ggtgaggctg gggggaggag cttatgataa agttccctta 4671
 agaaactaga aaataaagat gaaaacatag aattaagggt tttttgtttc tttcttcctt 4731
 50 tttttttttt ttttgtacta agaaataaga ttgaacagtg gatactgaaa tttggtgaat 4791
 tattttgtaa gtgattctct catttgtctt tctgaagcta cagctgttca tcatcact 4851
 acccttacc tgtctatcca ttctgtcatt gtcacaaaa aaaaaaagtc agtaattact 4911
 55 agctacaaaa ctatctaaca agcccttctc tggatgattt actttgtgtt aaagacttac 4971
 acagatttat aatcacattt agttgtgtgg cattaccaca atatgactca aagcaaaagc 5031
 60 agacttctgt ctgtttagt gtttttaagt gtgtgttgtg ggggtgggga gggsrsdbac 5091
 k 5092

<210> 4
 <211> 1207
 <212> PRT
 <213> Mus sp.

5

<220>
 <221> Amino acid sequence encoding Mouse NCAG1 protein

<400> 4
 10 Met Ala Phe Met Phe Thr Glu His Leu Leu Phe Leu Thr Leu Met Met
 1 5 10 15
 Cys Ser Phe Ser Thr Cys Glu Glu Ser Val Ser Asn Tyr Ser Glu Trp
 20 25 30
 15 Ala Val Phe Thr Asp Asp Ile Gln Trp Leu Lys Ser Gln Lys Ile Gln
 35 40 45
 Asp Phe Lys Leu Asn Arg Arg Leu His Pro Asn Leu Tyr Phe Asp Ala
 20 50 55 60
 Gly Asp Ile Gln Thr Leu Lys Gln Lys Ser Arg Thr Ser His Leu His
 65 70 75 80
 25 Ile Phe Arg Ala Ile Lys Ser Ala Val Thr Ile Met Leu Ser Asn Pro
 85 90 95
 Ser Tyr Tyr Leu Pro Pro Pro Lys His Ala Glu Phe Ala Ala Lys Trp
 100 105 110
 30 Asn Glu Ile Tyr Gly Asn Asn Leu Pro Pro Leu Ala Leu Tyr Cys Leu
 115 120 125
 Leu Cys Pro Glu Asp Lys Val Ala Phe Glu Phe Val Met Glu Tyr Met
 35 130 135 140
 Asp Arg Met Val Ser Tyr Lys Asp Trp Leu Val Glu Asn Ala Pro Gly
 145 150 155 160
 40 Asp Glu Val Pro Val Gly His Ser Leu Thr Gly Phe Ala Thr Ala Phe
 165 170 175
 Asp Phe Leu Tyr Asn Leu Leu Gly Asn Gln Arg Lys Gln Lys Tyr Leu
 180 185 190
 45 Glu Lys Ile Trp Ile Val Thr Glu Glu Met Tyr Glu Tyr Ser Lys Ile
 195 200 205
 Arg Ser Trp Gly Lys Gln Leu Leu His Asn His Gln Ala Thr Asn Met
 50 210 215 220
 Ile Ala Leu Leu Ile Gly Ala Leu Val Thr Gly Val Asp Lys Gly Ser
 225 230 235 240
 55 Lys Ala Asn Ile Trp Lys Gln Val Val Val Asp Val Met Glu Lys Thr
 245 250 255
 Met Phe Leu Leu Lys His Ile Val Asp Gly Ser Leu Asp Glu Gly Val
 260 265 270
 60 Ala Tyr Gly Ser Tyr Thr Ser Lys Ser Val Thr Gln Tyr Val Phe Leu
 275 280 285

	Ala	Gln	Arg	His	Phe	Asn	Ile	Asn	Asn	Phe	Asp	Asn	Asn	Trp	Leu	Lys	
	290						295					300					
5	Met	His	Phe	Trp	Phe	Tyr	Tyr	Ala	Thr	Leu	Leu	Pro	Gly	Tyr	Gln	Arg	
	305					310					315					320	
	Thr	Val	Gly	Ile	Ala	Asp	Ser	Asn	Tyr	Asn	Trp	Phe	Tyr	Gly	Pro	Glu	
					325					330					335		
10	Ser	Gln	Leu	Val	Phe	Leu	Asp	Lys	Phe	Ile	Leu	Gln	Asn	Gly	Ala	Gly	
				340					345					350			
	Asn	Trp	Leu	Ala	Gln	Gln	Ile	Arg	Lys	His	Arg	Pro	Lys	Asp	Gly	Pro	
15			355					360					365				
	Met	Val	Pro	Ser	Thr	Ala	Gln	Arg	Trp	Ser	Thr	Leu	His	Thr	Glu	Tyr	
		370					375					380					
20	Ile	Trp	Tyr	Asp	Pro	Thr	Leu	Thr	Pro	Gln	Pro	Pro	Val	Asp	Phe	Gly	
	385					390					395					400	
	Thr	Ala	Lys	Met	His	Thr	Phe	Pro	Asn	Trp	Gly	Val	Val	Thr	Tyr	Gly	
					405					410					415		
25	Gly	Gly	Leu	Pro	Asn	Thr	Gln	Thr	Asn	Thr	Phe	Val	Ser	Phe	Lys	Ser	
				420					425					430			
	Gly	Lys	Leu	Gly	Gly	Arg	Ala	Val	Tyr	Asp	Ile	Val	His	Phe	Gln	Pro	
30			435					440					445				
	Tyr	Ser	Trp	Ile	Asp	Gly	Trp	Arg	Ser	Phe	Asn	Pro	Gly	His	Glu	His	
		450				455						460					
35	Pro	Asp	Gln	Asn	Ser	Phe	Thr	Phe	Ala	Pro	Asn	Gly	Gln	Val	Phe	Val	
	465					470					475					480	
	Ser	Glu	Ala	Leu	Tyr	Gly	Pro	Lys	Leu	Ser	His	Leu	Asn	Asn	Val	Leu	
					485					490					495		
40	Val	Phe	Ala	Pro	Ser	Pro	Ser	Ser	Gln	Cys	Asn	Gln	Pro	Trp	Glu	Gly	
				500					505					510			
	Gln	Leu	Gly	Glu	Cys	Ala	Gln	Trp	Leu	Lys	Trp	Thr	Gly	Glu	Glu	Val	
45			515					520					525				
	Gly	Asp	Ala	Ala	Gly	Glu	Val	Ile	Thr	Ala	Ala	Gln	His	Gly	Asp	Arg	
		530				535						540					
50	Met	Phe	Val	Ser	Gly	Glu	Ala	Val	Ser	Ala	Tyr	Ser	Ser	Ala	Met	Arg	
	545					550					555					560	
	Leu	Lys	Ser	Val	Tyr	Arg	Ala	Leu	Leu	Leu	Leu	Asn	Ser	Gln	Thr	Leu	
					565					570					575		
55	Leu	Val	Val	Asp	His	Ile	Glu	Arg	Gln	Glu	Thr	Ser	Pro	Ile	Asn	Ser	
				580					585					590			
	Val	Ser	Ala	Phe	Phe	His	Asn	Leu	Asp	Ile	Asp	Phe	Lys	Tyr	Ile	Pro	
60			595					600					605				
	Tyr	Lys	Phe	Met	Asn	Arg	Tyr	Asn	Gly	Ala	Met	Met	Asp	Val	Trp	Asp	
		610					615					620					

	Ala	His	Tyr	Lys	Met	Phe	Trp	Phe	Asp	His	His	Gly	Asn	Ser	Pro	Val	
	625					630					635					640	
5	Ala	Asn	Ile	Gln	Glu	Ala	Glu	Gln	Ala	Ala	Glu	Phe	Lys	Lys	Arg	Trp	
				645					650						655		
	Thr	Gln	Phe	Val	Asn	Val	Thr	Phe	His	Met	Glu	Ser	Thr	Ile	Thr	Arg	
				660				665						670			
10	Ile	Ala	Tyr	Val	Phe	Tyr	Gly	Pro	Tyr	Val	Asn	Val	Ser	Ser	Cys	Arg	
			675				680						685				
	Phe	Ile	Asp	Ser	Ser	Ser	Ser	Gly	Leu	Gln	Ile	Ser	Leu	His	Val	Asn	
15		690					695					700					
	Ser	Thr	Glu	His	Ser	Val	Ser	Val	Val	Thr	Asp	Tyr	Gln	Asn	Leu	Lys	
	705					710					715					720	
20	Ser	Arg	Phe	Ser	Tyr	Leu	Gly	Phe	Gly	Gly	Phe	Ala	Ser	Val	Ala	Asn	
					725					730					735		
	Gln	Gly	Gln	Ile	Thr	Arg	Phe	Gly	Leu	Gly	Thr	Gln	Glu	Ile	Val	Asn	
				740					745					750			
25	Pro	Val	Arg	His	Asp	Lys	Val	Asn	Phe	Pro	Phe	Gly	Phe	Lys	Phe	Asn	
			755					760					765				
	Ile	Ala	Val	Gly	Phe	Ile	Leu	Cys	Ile	Ser	Leu	Val	Ile	Leu	Thr	Phe	
30		770					775					780					
	Gln	Trp	Arg	Phe	Tyr	Leu	Ser	Phe	Arg	Lys	Leu	Met	Arg	Cys	Val	Leu	
	785					790					795					800	
35	Ile	Leu	Val	Ile	Ala	Leu	Trp	Phe	Ile	Glu	Leu	Leu	Asp	Val	Trp	Ser	
					805					810					815		
	Thr	Cys	Thr	Gln	Pro	Ile	Cys	Ala	Lys	Trp	Thr	Arg	Thr	Glu	Ala	Lys	
				820					825					830			
40	Ala	Asn	Glu	Lys	Val	Met	Ile	Ser	Glu	Gly	His	His	Val	Asp	Leu	Pro	
			835					840					845				
	Asn	Val	Ile	Ile	Thr	Ser	Leu	Pro	Gly	Ser	Gly	Ala	Glu	Ile	Leu	Lys	
45		850					855					860					
	Gln	Leu	Phe	Phe	Asn	Ser	Ser	Asp	Phe	Leu	Tyr	Ile	Arg	Ile	Pro	Thr	
	865					870					875					880	
50	Ala	Tyr	Met	Asp	Ile	Pro	Glu	Thr	Glu	Phe	Glu	Ile	Asp	Ser	Phe	Val	
					885					890					895		
	Asp	Ala	Cys	Glu	Trp	Lys	Val	Ser	Asp	Ile	Arg	Ser	Gly	His	Phe	His	
				900					905					910			
55	Leu	Leu	Arg	Gly	Trp	Leu	Gln	Ser	Leu	Val	Gln	Asp	Thr	Lys	Leu	His	
			915				920						925				
	Leu	Gln	Asn	Ile	His	Leu	His	Glu	Thr	Ser	Arg	Ser	Lys	Leu	Ala	Gln	
60		930					935					940					
	Tyr	Phe	Thr	Thr	Asn	Lys	Asp	Lys	Lys	Arg	Lys	Leu	Lys	Arg	Arg	Glu	
	945					950					955					960	

Ser Leu Gln Asp Gln Arg Ser Arg Ile Lys Gly Pro Phe Asp Arg Asp
 965 970 975
 5 Ala Glu Tyr Ile Arg Ala Leu Arg Arg His Leu Val Tyr Tyr Pro Ser
 980 985 990
 Ala Arg Pro Val Leu Ser Leu Ser Ser Gly Ser Trp Thr Leu Lys Leu
 995 1000 1005
 10 His Phe Phe Gln Glu Val Leu Gly Thr Ser Met Arg Ala Leu Tyr Ile
 1010 1015 1020
 Val Arg Asp Pro Arg Ala Trp Ile Tyr Ser Val Leu Tyr Gly Ser Lys
 025 1030 1035 1040
 15 Pro Ser Leu Tyr Ser Leu Lys Asn Val Pro Glu His Leu Ala Lys Leu
 1045 1050 1055
 20 Phe Lys Ile Glu Glu Gly Lys Ser Lys Cys Asn Ser Asn Ser Gly Tyr
 1060 1065 1070
 Ala Phe Glu Tyr Glu Ser Leu Lys Lys Glu Leu Glu Ile Ser Gln Ser
 1075 1080 1085
 25 Asn Ala Ile Ser Leu Leu Ser His Leu Trp Val Ala Asn Thr Ala Ala
 1090 1095 1100
 Ala Leu Arg Ile Asn Thr Asp Leu Leu Pro Thr Asn Tyr His Leu Val
 105 1110 1115 1120
 30 Lys Phe Glu Asp Ile Val His Phe Pro Gln Lys Thr Thr Glu Arg Ile
 1125 1130 1135
 35 Phe Ala Phe Leu Gly Ile Pro Leu Ser Pro Ala Ser Leu Asn Gln Met
 1140 1145 1150
 Leu Phe Ala Thr Ser Thr Asn Leu Phe Tyr Leu Pro Tyr Glu Gly Glu
 1155 1160 1165
 40 Ile Ser Pro Ser Asn Thr Asn Ile Trp Lys Thr Asn Leu Pro Arg Asp
 1170 1175 1180
 Glu Ile Lys Leu Ile Glu Asn Ile Cys Trp Thr Leu Met Asp His Leu
 185 1190 1195 1200
 45 Gly Tyr Pro Lys Phe Met Asp
 1205